МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ

ХАРКІВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ МІСЬКОГО ГОСПОДАРСТВА імені О. М. БЕКЕТОВА

О. О. ВОРОНКОВ
І. А. МЕДВЕДЄВ

КОНСПЕКТ ЛЕКЦІЙ
з курсу

ЕКОНОМІКО-МАТЕМАТИЧНІ МЕТОДИ І МОДЕЛІ

(для слухачів другої вищої освіти за спеціальностями 7.03050401 – Економіка підприємства, 7.03050901 – Облік і аудит)

Харків
ХНУМГ ім. О. М. Бекетова
2016

Воронков О. О. Конспект лекцій з курсу «Економіко-математичні методи і моделі» (для слухачів другої вищої освіти за спеціальностями 7.03050401 – Економіка підприємства, 7.03050901 – Облік і аудит) / О. О. Воронков, І. А. Медведєв; Харків. нац. ун-т міськ. госп-ва ім. О. М. Бекетова. – Харків : ХНУМГ ім. О. М. Бекетова, 2016. – 199 с.

Автори: канд. екон. наук О. О. Воронков
канд. наук з держ. упр. І. А. Медведєв

Рецензент: канд. екон. наук, доцент Н. І. Склярук

Рекомендовано кафедрою економіки підприємств міського господарства, протокол засідання № 1 від 29 серпня 2014 р.

ЗМІСТ

ВСТУП

Курс «Економіко-математичні методи і моделі» є нормативною дисципліною циклу природничо-наукової і загальноекономічної підготовки в навчальному плані спеціалістів за спеціальностями «Економіка підприємства» та «Облік і аудит. Обсяг дисципліни для слухачів другої вищої освіти становить 144 академічних години або 4 кредити. При вивченні за заочною формою обсяг аудиторних занять становить 16 годин (8 годин лекцій і 8 годин практичних занять), на самостійну роботу студента припадає 128 години. Програма курсу містить 3 змістових модулі: «Оптимізаційні економіко-математичні моделі», «Аналіз та управління ризиком в економіці» і «Економетричні моделі», відповідно до яких виконується поточний контроль знань. Підсумковий контроль знань (іспит) проводиться у письмовій формі. Під час вивчення курсу студенти мають виконати контрольну роботу.

Метою вивчення дисципліни «Економіко-математичні методи і моделі» є формування системи базових знань в області методології постановки задач, побудови економіко-математичних моделей і методів їх розв'язання і аналізу.

В результаті вивчення курсу студенти повинні оволодіти прийомами побудови економіко-математичних моделей, основними математичними поняттями і методами розв'язання оптимізаційних задач різної складності, елементами теорії ігор, теорії імовірностей і математичної статистики, а також економетрії.

Сучасна економічна наука характеризується широким використанням математики. Математичні методи є складовою методів будь-якої економічної науки, включаючи економічну теорію. Використання математичних методів відчиняє нові можливості, і фахівцеві необхідно вміти формулювати й розв'язувати задачі з оптимізації виробництва, моделювати економічну динаміку і ризикові ситуації, статистично оцінювати економічні залежності, а також користуватися ігровими методами.

Математичні моделі використовувалися з ілюстративними і дослідницькими цілями ще Ф. Кене (1758 р., "Економічна таблиця"), А. Смітом (класична макроекономічна модель), Д. Рикардо (модель міжнародної торгівлі). В XIX столітті великий внесок у моделювання ринкової економіки внесла математична школа (Л. Вальрас, О. Курно, В. Парето, Ф. Еджворт та ін.). В XX столітті математичні методи моделювання застосовувалися дуже широко, з їх використанням пов'язані практично всі роботи, що відзначені Нобелівською премією з економіки (Д. Хикс, Р. Солоу, В. Леонтьєв, П. Самуельсон та ін.). Розвиток мікроекономіки, макроекономіки, прикладних дисциплін пов'язаний з усе більш високим рівнем їх формалізації. Основу для цього заклав прогрес в області прикладної математики - теорії ігор, математичного програмування, математичної статистики.

Перші роботи з економетрики з'явилися наприкінці XIX – початку XX століття. В 1897 р. з'явилася робота одного із засновників математичної школи в економічній теорії В. Парето, яка присвячена статистичному вивченню доходів населення в різних країнах. Було запропоновано криву Парето $y = A(x-a)^{-\alpha}$, де x – величина доходу; y – чисельність осіб, що мають дохід, що перевищує x; a – мінімальний дохід; A і α – параметри залежності, одержувані за статистичними методами.

На самому початку XX століття вийшло кілька робіт англійського статистика Гукера, у яких він застосував кореляційно-регресійні методи, розроблені Пірсоном і його школою, для вивчення взаємозв'язків економічних показників, зокрема – впливу числа банкрутств на товарній біржі на ціну зерна. У роботах Гукера містилася ідея часового лага між економічними змінними, а також ідея кореляційного аналізу не самих величин, а їх приростів. Надалі з'явилося величезне число робіт як з розвитку теорії математичної статистики та її прикладних елементів, так і з практичного застосування цих методів в економічному аналізі. До першої групи можуть бути, наприклад, віднесені роботи Р. Фішера з дисперсійного аналізу, до другої - роботи з оцінки й дослідження виробничих функцій, зокрема – класична робота Кобба й Дугласа 1928 р. Економетричні моделі й методи зараз – це не тільки могутній інструментарій для одержання нових знань в економіці, але й широко застосовуваний апарат для прийняття практичних рішень у прогнозуванні, банківській справі, бізнесі.

ТЕМА 1
КОНЦЕПТУАЛЬНІ АСПЕКТИ МАТЕМАТИЧНОГО МОДЕЛЮВАННЯ ЕКОНОМІКИ

1.1 Економічні моделі. Поняття економічної моделі

Сучасна економічна теорія як на мікро-, так і на макрорівні, включає як необхідний елемент математичні моделі і методи. Використання математики в економіці дозволяє, по-перше, виділити і формально описати найбільш важливі, істотні зв'язки економічних змінних і об'єктів. По-друге, із чітко сформульованих вихідних даних і співвідношень за методами дедукції можна одержувати висновки, адекватні досліджуваному об'єкту. По-третє, методи математики і статистики дозволяють індуктивним шляхом одержувати нові знання про об'єкт: оцінювати форму і параметри залежностей між його змінними, що найбільшою мірою відповідають наявним спостереженням. І, нарешті, по-четверте, використання математичної мови дозволяє точно і компактно викладати положення економічної теорії, формулювати її поняття і висновки.

Для вивчення різних економічних явищ економісти використовують їх спрощені формальні описи, називані **економічними моделями**. Прикладами економічних моделей є моделі споживчого вибору, моделі фірми, моделі економічного зростання, моделі рівноваги на товарних, факторних і фінансових ринках і багато інших. При побудові моделі виявляють істотні фактори, що зумовлюють досліджуване явище і відкидають деталі, які є несуттєвими для розв'язання поставленої проблеми. Формалізація основних особливостей функціонування економічних об'єктів дозволяє оцінити можливі наслідки впливу на них і використовувати такі оцінки в управлінні.

Модель (лат. – міра, зразок) – це певний об'єкт, що у певних умовах заміняє об'єкт-оригінал, відтворюючи властивості, які цікавлять дослідника, і характеристики оригіналу, маючи при цьому істотні переваги використання (наочність, доступність випробувань та ін.).

Моделювання – це дослідження будь-яких процесів, явищ або систем (об'єктів) шляхом побудови й вивчення їх моделей; використання моделей для визначення або уточнення характеристик і раціоналізації способів побудови новостворюваних об'єктів.

Моделювання – одна з основних категорій теорії пізнання. На ідеї моделювання базується будь-який метод наукового дослідження, як теоретичний (при якому використовуються абстрактні моделі), так і експериментальний (що використовує предметні моделі).

1.2 Принципи моделювання

Побудова і використання моделей базується на наступних основних принципах.

1. Принцип інформаційної достатності. При повній відсутності інформації про систему побудова її моделі є неможливою. При наявності повної інформації про систему відсутня необхідність її моделювання. У всіх проміжних ситуаціях дослідник змушений оперувати з моделлю системи, рівень адекватності якої зумовлений певним критичним рівнем апріорної інформації про систему.

2. Принцип здійсненності. Модель, розроблювальна дослідником, повинна забезпечити досягнення поставленої мети з практичною вірогідністю і за кінцевий час.

3. Принцип множинності моделей. Складність досліджуваних систем і велика розмаїтість їх властивостей не дозволяють побудувати одну досить адекватну модель. Тому виникає необхідність у побудові кількох моделей, які в сукупності дають досить повне уявлення про систему і всі процеси, що протікають у ній.

Будь-яка діяльність людини носить цільовий характер. Моделювання є неминучою процедурою у всякій доцільній діяльності. Треба особливо підкреслити, що модель є не взагалі якимсь відображенням оригіналу, а цільовим відображенням. Ціль моделювання обумовлює, які властивості оригіналу і якою мірою (з якою точністю) повинні бути відображені в моделі.

Зазвичай під час побудови економічної моделі проходять наступні етапи.

1. Формулюють предмет і цілі дослідження.

2. У розглянутій економічній системі виділяють структурні або функціональні елементи, що відповідають певній цілі, виявляють найбільш важливі якісні характеристики цих елементів.

3. Словесно, якісно описують взаємозв'язки між елементами моделі.

4. Вводять символічні позначення для характеристик економічного об'єкта, що враховують у моделі, і формалізують, наскільки можливо, взаємозв'язки між ними. Тим самим, формулюється математична модель.

5. Проводять розрахунки за математичною моделлю і виконують аналіз отриманого розв'язку.

Розглянемо математичну структуру моделі та її змістову інтерпретацію.

Приклад 1. Нехай потрібно визначити, яку суму треба покласти в банк при заданій ставці відсотка (20% річних), щоб через рік одержати \$12000. Введемо формальні позначення для величин: початкова сума грошей – M_0; кінцева сума грошей – M_1; ставка відсотка – R, запишемо співвідношення між ними

$$M_1 = M_0\left(1 + \frac{R}{100}\right),$$

і знайдемо необхідну величину з розв'язку основного рівняння моделі

$$M_0 = \frac{M_1}{1 + \frac{R}{100}} = \frac{\$12000}{1{,}2} = \$10000\,.$$

Приклад 2. Нехай потрібно визначити, який був обсяг випуску продукції заводу, якщо в результаті технічного переозброєння середня продуктивність праці збільшилася на 20%, і завод став випускати 12000 одиниць продукції. Введемо формальні позначення для величин: початковий випуск – Q_0; кінцевий випуск – Q_1; відсоток приросту продуктивності – R, запишемо співвідношення між ними (що випливає з визначення середньої продуктивності праці $\frac{Q}{L}$)

$$Q_1 = Q_0\frac{L_1}{L_0} = Q_0\left(1 + \frac{L_1 - L_0}{L_0}\right) = Q_0\left(1 + \frac{R}{100}\right),$$

та знайдемо шукану величину з розв'язку основного рівняння моделі

$$Q_0 = \frac{Q_1}{1 + \frac{R}{100}} = \frac{\$12000}{1{,}2} = \$10000\,.$$

З порівняння отриманих моделей і результатів помітимо, що математична форма моделі та сама

$$X_1 = X_0\left(1 + \frac{R}{100}\right),$$

і навіть числові значення величин, що входять до неї, в обох випадках однакові, але економічна ситуація, описувана моделлю, економічна інтерпретація моделі і результатів розрахунку зовсім різні. Отже, ті самі математичні моделі і методи можуть бути використані для розв’язання зовсім різних економічних завдань.

1.3 Класифікація моделей

Підставою для класифікації моделей є ціль моделювання. Цільова орієнтація моделей дозволяє класифікувати їх за типами цілей, за способами відтворення (реалізації), за зміною на різних етапах життєвого циклу.

За типами цілей розрізняють моделі пізнавальні і прагматичні. За способами відтворення розрізняють моделі ідеальні (абстрактні) і матеріальні (реальні, фізичні). За зміною в часі розрізняють моделі статичні і динамічні.

Розходження пізнавальних і прагматичних моделей проявляється в їх ставленні до оригіналу в процесі діяльності. **Пізнавальні моделі** є формою організації і подання знань, засобами з'єднання нових знань з наявними. Якщо в процесі створення пізнавальної моделі певного реального об'єкта спостерігаються розбіжності, то здійснюється корекція моделі з метою наблизити її до реальності.

Прагматичні моделі є засобами управління практичними діями, способом уявлення необхідних дій або їх результату.

Прагматичні і пізнавальні моделі можуть створюватися як для певного фіксованого моменту часу, тобто бути статичними, так і для певного інтервалу життєвого циклу об'єкта моделювання в умовах змінного середовища. У цьому випадку необхідною є динамічна модель, параметри (характеристики) якої залежать від часу. Динамічні моделі зазвичай значно складніше за статичні, вимагають застосування складного математичного апарата і великого обсягу машинної пам'яті. Для спрощення динамічні моделі часто розглядають у вигляді набору послідовностей статичних моделей, що відповідають певним моментам часу на траєкторії життєвого циклу об'єкта.

Ідеальними називають моделі, що побудовані засобами мислення, свідомості. До цих моделей належать всі язикові конструкції, що сприяють установленню відносин між людьми.

Природна мова є універсальним засобом побудови абстрактних конструкцій (моделей). Ця універсальність забезпечується можливістю введення в мову нових слів і можливістю ієрархічної побудови усе більш розвинених язикових моделей. Крім цього, універсальність мови досягається тим, що язикові конструкції мають неоднозначність, розпливчастість, розмитість. Людина переборює цю неоднозначність за допомогою «розуміння» й «інтерпретації», що залежить від культурного середовища, у якому віна живе.

Якщо на побутовому рівні приблизність мови сприяє поліпшенню відносин серед людей, то при вирішенні ряду практичних і наукових завдань це стає перешкодою, наприклад, у діалозі з ЕОМ. Тому з'являється «професійна мова», реалізована в обмежених професійних колективах, мова математиків, фізиків, медиків, льотчиків, фахівців з інформатики та ін.

Найбільш високо спеціалізованою є мова математики, що має максимально досяжну на сьогодні визначеність і точність.

Еммануїл Кант відзначав, що «у кожному пізнанні стільки науки, скільки є в ньому математики». Низький рівень математизації будь-якої науки не означає її «ненауковість», а просто є наслідком її складності, недостатньої глибини пізнання предмета науки й, отже, явищем тимчасовим. Наочним прикладом цьому є зміна поглядів людей на проблему створення «мислячої машини». Якщо в 50-х роках 20-го сторіччя розмови про це викликали тільки посмішки скептиків, то в цей час з'явився цілий напрямок у математиці - теорія штучного інтелекту, на підставі якої базуються високоефективні експертні системи.

За характером обчислення різних показників математичні моделі поділяються на аналітичні, алгоритмічні й імітаційні.

Аналітичні моделі передбачають реалізацію моделі у вигляді алгебраїчних, диференціальних, інтегральних і інших рівнянь, що зв'язують вихідні змінні із вхідними, доповнених системою обмежень. При цьому передбачають наявність однозначної обчислювальної процедури одержання точного розв'язку рівнянь.

При алгоритмічному підході використовувана математична модель не припускає точного подання у вигляді системи рівнянь, а залежність між вхідними (незалежними) і вихідними (залежними) змінними задається певною послідовністю дій. Такий підхід при створенні моделей складних систем є типовим.

Основним видом математичних моделей складних систем є імітаційні моделі.

Імітаційна модель являє собою певну обчислювальну процедуру, що описує об'єкт аналізу. Вона дозволяє з будь-якою заданою точністю відтворити параметричну систему будь-якої складності. Основними обмеженнями при створенні цих моделей є ресурси пам'яті й часу. Головними засобами реалізації імітаційних моделей є ЕОМ. Імітаційне моделювання, являючись своєрідним машинним експериментуванням з моделлю реальної системи, відчиняє широкі можливості для адекватного відображення процесу функціонування складних систем.

Математичні моделі, використовувані в економіці, можна підрозділити на класи за рядом ознак, що належать до особливостей об'єкта моделювання, цілі моделювання і використовуваного інструментарію: моделі макро- і мікроекономічні, теоретичні і прикладні, оптимізаційні і рівноважні, статичні і динамічні.

Теоретичні моделі дозволяють вивчати загальні властивості економіки та її характерних елементів шляхом дедукції висновків з формальних передумов. **Прикладні моделі** дають можливість оцінити параметри функціонування конкретного економічного об'єкта і сформулювати рекомендації для прийняття практичних рішень. До прикладних належать насамперед економетричні моделі, що оперують числовими значеннями економічних змінних і дозволяють статистично значущо оцінювати їх на основі наявних спостережень.

У моделях **статичних** описується стан економічного об'єкта в конкретний момент або період часу; **динамічні моделі** включають взаємозв'язки змінних у часі. У статичних моделях зазвичай зафіксовані значення ряду величин, що є змінними в динаміці, наприклад, капітальних ресурсів, цін та ін. Динамічна модель не зводиться до простої суми ряду статичних, вона описує сили і взаємодії в економіці, що обумовлюють протікання процесів у ній. Динамічні моделі зазвичай використовують апарат диференційних і різницевих рівнянь, а також варіаційного обчислення.

Детерміновані моделі припускають жорсткі функціональні зв'язки між змінними моделі. **Стохастичні моделі** припускають наявність випадкових впливів на досліджувані показники і використовують інструментарій теорії імовірностей і математичної статистики для їх опису.

1.4 Якість моделі

Якість моделі, як і будь-якого іншого об'єкта, є сукупністю різних властивостей, що мають ієрархічну структуру. На першому рівні виділяють три групи властивостей: цільові, експлуатаційні, модифікаційні.

Цільовими називають властивості моделі, що характеризують ступінь її відповідності цілям і завданням конкретного об'єкта моделювання. Цільові властивості включають наступні властивості другого рівня: адекватність, стійкість, точність, результативність.

Експлуатаційними називають властивості моделі системи, що визначають зручність експлуатації моделі користувачем. До них належать наступні властивості другого рівня: продуктивність, надійність, захищеність, завершеність.

Модифікаційними називають властивості моделі, що зумовлюють зручності внесення змін до моделі, простоту модернізації моделі при розвитку й удосконалюванні досліджуваної системи при переході на інший рівень дослідження (наприклад, при підвищенні ступеня деталізації математичного опису). До цієї групи належать наступні властивості: зрозумілість, структурованість, розширюваність, доступність.

У людській практиці спрощеність моделей є припустимою, більше того, у певних випадках вона є необхідною. І хоча спрощеність моделей призводить до розходження між моделлю і оригіналом, міру припустимого розходження можна оцінити тільки, якщо співвіднести її з ціллю моделювання. Тому для оцінки ступеня відповідності моделі й оригіналу вводять таке поняття як "**адекватність моделі**".

> *Адекватною називають таку модель, для якої вимоги повноти, точності й істинності моделі виконуються не взагалі повною мірою, а тільки в тому ступені, що забезпечує досягнення цілі.*

1.5 Прийняття рішень (вибір)

Вибір – це є дія, що дозволяє організувати цілеспрямовану діяльність людини, а рішення – результат цього вибору у певній нормативно-правовій формі (порада, рекомендація, наказ, програма та ін.).

Управляючи будь-якою діяльністю, людина постійно зіштовхується з необхідністю приймати рішення на ту або іншу дію. Отже, ухвалення рішення розглядають як процедуру вибору альтернативи із заданої множини.

Альтернатива (фр.) – кожна з взаємно виключаючих двох або більше можливостей.

Для того щоб прийняти рішення, необхідно виконати наступну послідовність процедур:

1) вибрати цілі, заради досягнення яких здійснюється вибір;

2) оцінити ступінь погодженості цілей (від повної погодженості до повного протиріччя);

3) сформулювати множину альтернатив, з яких здійснюється вибір;

4) проаналізувати і оцінити наслідки реалізації кожної альтернативи;

5) сформулювати критерії порівняння, тобто правило, за допомогою якого визначається переваги альтернатив;

6) задати режим вибору: однократний або багаторазовий з навчанням;

7) оцінити ситуацію, у якій здійснюється вибір (визначеність або невизначеність та її вид);

8) визначити тип відповідальності (індивідуальна, групова).

Сучасна наука розрізняє дві частини теорії прийняття рішень. Одна з них має нормативний характер і займається дослідженнями технологій прийняття рішень для того, щоб можна було відповістити на запитання: «Як необхідно приймати рішення?». Інша частина досліджує питання, пов'язані з тим, як люди на практиці приймають рішення, і які при цьому вони роблять помилки.

Нормативну теорію прийняття рішень у літературі часто називають теорією прийняття оптимального рішення, а другу частину - психологічною теорією рішень.

Теорія прийняття оптимального рішення розвивалася завдяки успіхам, досягнутим в області дослідження операцій і системотехніці. Ця теорія розробляє методологію прийняття рішень з метою організації діяльності з досягнення певної цілі, але не дає відповідь на питання про те, як ця ціль вибирається і чи відповідає вона бажаному сценарію розвитку подій.

Вона оперує з процедурами і критеріями прийняття рішень, які можна вважати оптимальними в рамках тієї моделі розглянутої ситуації, якою керувалася особа, що приймає рішення, опираючись на свій рівень знань і наявну інформацію.

Одним з найважливіших напрямків теорії прийняття оптимальних рішень є опис умов, які повинні бути виконані для того, щоб рішення було оптимальним. Ці умови формулюють у вигляді постулатів оптимальності, серед яких найбільше визнання одержали два: постулати послідовності і максимізації.

Постулат послідовності говорить, що для прийняття оптимального рішення треба впорядкувати сукупність альтернатив з погляду переваг особи, що приймає рішення (ОПР). Якщо розглядають три альтернативи A, B і C, то впорядкування цих альтернатив означає побудову шкали строгого або слабкого порядку альтернатив за ступенем убування їх переваги. Наприклад, порядок альтернатив виду $A > C > B$ означає, що при здійсненні вибору A переважніше C, а C переважніше B.

Постулат максимізації стверджує, що остаточною умовою оптимального рішення є використання принципу максимізації, тобто вибір такої дії, що максимізує цільову функцію.

Формалізація моделей прийняття рішень для розглянутих ситуацій здійснюється в термінах теорії імовірностей, математичної статистики, теорії масового обслуговування, теорії; ігор та ін.

Застосування формальних оптимізаційних методів підготовки і прийняття рішень можливо тільки в умовах досить повної інформації. Така ситуація виникає, як правило, на нижніх рівнях управлінської структури, де як об'єкти управління розглядають технічні системи або технологічні процеси.

Область застосування оптимізаційних методів сильно звужується, коли доводиться приймати рішення на верхніх рівнях управління, де об'єкти, події, явища і процеси слабо структуровані.

Психологічна теорія прийняття рішень - це система мотивованих стверджень, що розкривають внутрішній зміст діяльності людини в процесі підготовки і прийняття рішень. Практика показує, що поведінка людини в сильному ступені залежить від структури розв'язуваного завдання. Отже, одним з головних напрямків розвитку цієї теорії є класифікація завдань і опис їх інваріантних властивостей з погляду їх впливу на характер прийняття рішень різними групами людей.

У психологічній теорії прийняття рішень виділяють кілька напрямків, для яких формулюють ствердження і рекомендації.

До першого напрямку належать питання, пов'язані з моделями уявлення оцінюваної ситуації різними категоріями людей. Наприклад, психологи виявили, що більшість людей схильні до спрощення і недооцінки рівня напруженості ситуації і не дуже охоче формулюють різні альтернативи.

Другий напрямок охоплює питання, пов'язані з моделюванням процесів визначення цінності того або іншого варіанту дії, з погляду його корисності для досягнення цілі.

Третій напрямок займається дослідженням суб'єктивної оцінки імовірностей появи тих або інших факторів і їх рівнів, які визначають наслідки прийнятих рішень. Так, наприклад, виявлено, що люди переоцінюють імовірності рідких подій і одночасно недооцінюють міру можливості появи тих подій, об'єктивна імовірність яких є великою.

Четвертий напрямок вивчає стратегії вибору поведінки людини при ухваленні рішення. Розробляють моделі, що описують, як ОПР інтерпретують інформацію про корисність наслідків і імовірності їх настання, і які правила вибору при цьому використовуються. Учені виявили, що в простих завданнях, пов'язаних з ризиком, люди зазвичай вибирають стратегії, які максимізують суб'єктивно очікувану корисність, яку вони уявляють як лінійну комбінацію суб'єктивних імовірностей наслідків та їх корисностей.

П'ятий напрямок займається описом факторів, керуючих процесом прийняття рішення. Такі фактори враховують вплив навколишнього природного середовища, особові особливості ОПР, вплив соціальної групи та ін.

Основою методології досліджень у психологічній теорії прийняття рішень є лабораторний експеримент, що включає кілька етапів.

На першому етапі формулюють сукупність аксіом і стверджень щодо певних об'єктів, таких, наприклад, як ризик або перевага.

На другому етапі за допомогою формальних логічних міркувань експертів виводять нові ствердження про можливий результат експерименту.

На третьому етапі здійснюють експериментальну перевірку висловлених гіпотез.

У цей час, поряд з натурними експериментами, для вирішення зазначених вище завдань широко використовують машинний експеримент.

Сучасні інформаційні технології, швидкодія і великий обсяг пам'яті ЕОМ дозволяють розробляти імітаційні моделі практично будь-якої складності і одержувати необхідні рекомендації.

1.6 Методи прийняття рішень

У психології прийняття людиною рішення пов'язують з процесом антиципації. Антиципація – це психічний процес, що забезпечує можливість приймати рішення з певним часово-просторовим попередженням подій, з «забіганням уперед».

У практиці людської діяльності розрізняють наступні методи прийняття рішень.

Автоматичний метод, реалізований на рівні біологічних механізмів інстинктів і рефлексів.

Метод проб і помилок, коли здійснюється несвідомий (некерований) або свідомий (керований) перебір ситуацій і пошук найкращого варіанта (режим самонавчання).

Метод звертання до авторитетів. Цей метод має багато прикладних відтінків: від спроби ОПР ухилятися від відповідальності (посиланням на бога, вождя, начальника та ін.) до спроби обпертися на досягнення кращих розумів людства.

Математичний метод є найбільш формалізованим методом прийняття рішень і вимагає наявності наступних обов'язкових компонентів:

- перелік цілей;
- перелік альтернатив;
- методи для прогнозування наслідків цих альтернатив;
- будь-який метод присвоєння імовірностей цим наслідкам (якщо вони існують);
- система цінностей, схована у цілях, для визначення цінності наслідків;
- критерій рішення, включений до системи цінностей, і який вказує, як визначити найкращу альтернативу.

Таку модель прийняття рішень називають **теорією статистичних рішень**.

На жаль, на практиці лише незначна частина рішень приймається за цим методом. Це пов'язано зі слабкою структурованістю цілей, низьким рівнем формалізації альтернатив та відсутністю вірогідної інформації про закони розподілу імовірностей наслідків, що розглядаються.

Залежно від типу зв'язку між альтернативами та їх наслідками виникають певні умови, у яких доводиться приймати рішення. Вони визначаються як умови визначеності або невизначеності.

Ухвалення рішення в умовах визначеності характеризуються наявністю повної і вірогідної інформації про проблемну ситуацію, цілі, обмеження і наслідки прийнятих рішень. У таких завданнях заздалегідь відомі наслідки всіх альтернатив.

На практиці приймати рішення доводиться в умовах, коли інформація про множину факторів, що впливають на вибір стратегії поведінки, як правило, неповна. Це призводить до того, що кожній розглянутій стратегії поведінки (альтернативі) у загальному випадку відповідає кілька можливих наслідків, що істотно ускладнює вибір стратегії.

Розрізняють стохастичну і не стохастичну невизначеності.

Стохастична невизначеність - це такий стан рівня інформованості ОПР, при якому для кожної розглянутої альтернативи відомі всі можливі наслідки та імовірності їх реалізації, тобто закон розподілу імовірностей наслідків. Прийняття рішення в умовах стохастичної невизначеності призводить до того, що поява того або іншого обраного результату є подією випадковою. Отже, після ухвалення рішення ОПР з тією або іншою імовірністю повинна очікувати, що відбудеться не та подія, на яку вона розраховувала. Завдання такого типу в теорії прийняття рішень називають «**прийняттям рішень в умовах ризику**».

Ризик визначають як незапланований (небажаний) результат розглянутого виду діяльності організації або спосіб функціонування якогось об'єкта.

Як міру ризику розглядають або імовірність небажаного результату, або математичне сподівання міри збитку, отриманого при реалізації стратегії.

Природно, що при прийнятті рішень у таких умовах ОПР повинна прагнути мінімізувати ризик або, принаймні, контролювати його рівень. При цьому виникають різні можливості організації поведінки ОПР, тобто прийняття рішень за різними критеріями.

У тому випадку, коли інформація про закон розподілу імовірностей наслідків відсутня, ОПР змушена використовувати методи прийняття рішень, не пов'язані із стохастичною природою розглянутих процесів, наприклад, методи теорії ігор. Такий вид невизначеностей називають не стохастичною невизначеністю.

Не стохастичну невизначеність підрозділяють на природну і поведінкову невизначеності. Перша пов'язана з нестачею інформації про об'єкти, явища і процеси, що відбуваються в природі, а друга - з поведінкою людини (або ОПР, або того, хто їй протидіє).

Поділ невизначеностей на природні і поведінкові носить штучний характер і враховує розходження в поведінці природи і людини при аналізі можливих варіантів протидії з метою активного впливу на кінцевий результат.

Якщо дослідник прагне оцінити і урахувати вплив природних факторів на прийняття рішення для досягнення бажаного результату, то він виходить з припущення, що природа є нейтральною стосовно дій людини, і найбільш розумним буде припущення про рівноможливості всіх розглянутих станів природи, тобто рівноможливості варіантів протидій при грі з природою.

При аналізі поведінкової невизначеності, тобто при грі з людиною, логічніше припустити, що свою поведінку супротивник (як фактор) організує найгіршим для нас (найкращим для себе) способом; і це треба враховувати при виборі критерію прийняття рішення.

Контрольні запитання

1. Чому необхідне використання математики в економіці?
2. Дайте визначення поняттю «модель».
3. Поясніть, що таке моделювання?
4. Поясніть, що таке математична модель?
5. Як будують математичну модель економічного явища або об'єкта? На-

ведіть приклад побудови й уточнення моделі.

6. Перелічіть і поясніть основні принципи моделювання.

7. Який зв'язок між моделлю і ціллю системи?

8. У чому відмінність статичних моделей від динамічних?

9. Що таке прагматична модель? Наведіть кілька прикладів практичного застосування таких моделей.

10. Дайте визначення ідеальним моделям. Наведіть приклади таких моделей.

11. Перелічіть й охарактеризуйте основні властивості якості моделі.

12. Що таке «адекватна модель»?

13. Дайте визначення поняттям «рішення» і «прийняття рішення».

14. Сформулюйте послідовність процедур, які необхідно виконати для прийняття рішення. Чи можна змінити цю послідовність? Які з процедур можуть виконуватися паралельно?

15. Охарактеризуйте дві частини теорії прийняття рішень і перелічіть основні завдання, які вони вирішують.

16. Сформулюйте основні постулати теорії прийняття оптимальних рішень.

17. Дайте характеристику основним видам невизначеностей, що виникають у процесі прийняття рішень.

18. Охарактеризуйте основні напрямки психологічної теорії прийняття рішень.

19. Перелічіть основні методи прийняття рішень і сформулюйте ситуації, у яких ці методи можна реалізувати.

ЗМ 1 Оптимізаційні економіко-математичні моделі

ТЕМА 2
ПОНЯТТЯ ОПТИМІЗАЦІЙНИХ ЗАДАЧ І ОПТИМІЗАЦІЙНИХ МОДЕЛЕЙ. КЛАСИФІКАЦІЯ

Економіко-математичні задачі, ціль яких полягає в знаходженні найкращого (оптимального) з погляду певного критерію або критеріїв варіанта використання наявних ресурсів (праці, капіталу та ін.), називають оптимізаційними.

Оптимізаційні задачі вирішують за допомогою оптимізаційних моделей за методами математичного програмування.

Структура оптимізаційної моделі складається з цільової функції, області припустимих рішень і системи обмежень, що визначають цю область. Цільова функція в самому загальному вигляді, у свою чергу, також складається з трьох елементів: керованих змінних, некерованих змінних, форми функції (виду залежності між ними).

Математичне програмування розробляє математичні методи кількісного обґрунтування рішень у задачах, пов'язаних з управлінням організаційними системами.

Нагадаємо, що **системою** називають сукупність взаємозв'язаних елементів, об'єднаних **загальною ціллю** функціонування. Під організаційною системою розуміють певну організацію, наприклад, виробниче підприємство. Хоча в загальному випадку природа таких систем може бути різною, зокрема, це можуть бути біологічні, соціальні або інші процеси.

Системи, що зустрічаються на практиці, в залежності від їх структури і характеру зв'язків поділяють на детерміновані і імовірнісні. **Детермінованою** називають систему, закони руху якої точно відомі й майбутню поведінку якої можна передбачити. Для **імовірнісної** системи не можна зробити точного прогнозування її майбутньої поведінки. Прикладом детермінованої системи може бути годинниковий механізм. Системи статистичного контролю продукції, системи прибуття кораблів у морські порти або запас товарів на складі, що має велику кількість постачальників і споживачів, є імовірнісними системами.

Ціллю системи називають певний (заданий ззовні або встановлюваний самою системою) найкращий кінцевий стан (наприклад, параметри вихідних характеристик), тобто певна підмножина значень функції системи.

Структура системи зумовлюється розташуванням і взаємозв'язками між складовими елементами системи, які утворені для виконання системою своєї функції, тобто залежить від величини й складності системи. Величина системи характеризується числом її елементів і кількістю зв'язків між ними, а складність - різноманіттям елементів, неоднорідністю їх властивостей і різною якістю зв'язків.

Великим і складним системам, таким як промислове підприємство, притаманні властивості **цілісності** й **емерджентності**.

Цілісність системи означає, що всі її елементи слугують спільній цілі й сприяють формуванню найкращих (оптимальних) результатів у смислі прийнятого критерію ефективності.

Емерджентність (поява нового) означає, що великі й складні системи мають такі властивості, які не властиві жодному з її елементів. Емерджентні або системні якості кардинально відрізняють системи від не систем. Інакше кажучи, об'єднання підсистем з різною природою (наприклад, економічної, технічної, соціальної) і структурою в складні системи відбувається при їх взаємовпливі одна на одну, що й створює нову системну якість, яка не властива ні однієї з підсистем.

Визначимо зміст терміна «**управління**». У широкому значенні слова під управлінням розуміють організаційну діяльність, що здійснює функції керівництва роботою, спрямованою на досягнення певних цілей. Процес управління полягає в **прийнятті рішень** про найбільш доцільні дії в тій або іншій сформованій ситуації. Людина, що здійснює управління, приймає рішення, оцінюючи навколишню обстановку за допомогою інформації, одержуваної від своїх органів почуттів, вимірювальних приладів, інших осіб. У багатьох випадках цієї інформації недостатньо для однозначної оцінки обстановки. Тоді людина використовує свій досвід, свої знання, пам'ять, інтуїцію. Чудовою властивістю людини є здатність приймати рішення в умовах значної невизначеності відносно навколишньої обстановки. Однак в умовах сучасних великих промислових підприємств знань і інтуїції навіть у досвідченого керівника недостатньо, щоб здійснювати ефективне управління. В результаті використовують математичні методи, у тому числі методи математичного програмування.

Управлінням називають цілеспрямований вплив на систему з метою отримати найкращий ефект. Можна сказати, що управління є такою організацією того або іншого процесу, що забезпечує досягнення певних цілей.

Можна виділити чотири етапи, характерні для будь-якого процесу управління: поява цілі, оцінка ситуації, ухвалення рішення і виконання ухваленого рішення. Оскільки етап появи цілі передує початку процесу управління, його можна виключити з розгляду. Врахуємо також, що при управлінні складними процесами оцінка ситуації провадиться на основі зібраної і обробленої інформації, таким чином, дістанемо наступні три етапи процесу управління: збір і обробка інформації з метою оцінки сформованої ситуації; ухвалення рішення про найбільш доцільні дії; виконання ухваленого рішення.

Управління завжди вимагає визначення певної ознаки (або критерію), що дозволяє оцінити ефективність прийнятого рішення, і який пояснює ступінь відповідності поведінки системи цілям управління.

Помітимо, що визначень ефективності багато. Часто ефективність визначають як умову, при якої система збільшить доходи при постійних або менших витратах або матиме постійний прибуток при менших витратах та ін. Загалом, ефективність можна уявити як кращий (необхідний) результат функціонування системи за менший час, при витраті менших ресурсів у довгостроковій і контрольованій перспективі. Уявлення про результат (ціль) має принципові відзнаки. Наприклад, переваги виробничої системи з вкладення ресурсів у нове обладнання, у навчання персоналу, або у збільшення жалування й розвиток соціальної сфери, або підвищення виплат власникам мають різну ефективність із позицій персоналу, власника, менеджера й стороннього спостерігача. Отже, визначення ефективності пов'язане з системою виражених і прихованих уявлень про її сутність, цілі, місце розташування спостерігача. Тому слід зазначити виняткову складність задачі строгої формалізації критеріїв ефективності.

У багатьох випадках реалізація процесу управління вимагає витрати будь-яких ресурсів: витрат часу, матеріалів, палива, електроенергії. Отже, при виборі способу управління треба говорити не тільки про те, чи досягається поставлена ціль, але й про те, які ресурси доведеться затратити для досягнення цієї цілі. У цьому випадку завдання з управління полягає в тому, щоб з множини рішень, що забезпечують досягнення цілі, вибрати одне рішення, що вимагає найменшої витрати ресурсів. В інших випадках підставою для переваги одного способу управління перед іншим можуть служити інші вимоги: вартість обслуговування, надійність та ін.

Математичний вираз, що дає кількісну оцінку ступеня досягнення цілі управління, називається **критерієм ефективності** управління. Найкращим або оптимальним способом управління буде такий, при якому критерій ефективності досягає **мінімального** або **максимального** значення.

При розв'язанні задачі управління треба зважати на ту обставину, що управління будь-якою системою завжди піддане різного роду обмеженням, викликаним обмеженістю ресурсів, використовуваних при управлінні, або інших величин, які в силу фізичних особливостей тієї або іншої системи не можуть або не повинні перевищувати певних меж. Математично обмеження цього виду виражають у вигляді систем алгебраїчних рівнянь або нерівностей, що пов'язують змінні, які описують стан системи. Тому особливе місце в методах оптимізації займає критерій оптимізації з обмеженнями. Смисл тут полягає в тому, що відшукують не будь-яке рішення x^*, що обертає критерій ефективності на мінімум, а таке, що задовольняє системі обмежень. Неважко переконатися, що значення умовного екстремуму не може бути меншим за значення абсолютного екстремуму (без обмежень). Задачі оптимізації при наявності обмежень призвели до перегляду класичних методів і створенню нових методів, відомих за назвою **методів програмування**.

Для визначення оптимальної стратегії при наявності конфліктної ситуації, коли інтереси двох сторін протилежні, широко використовують мінімаксний критерій. Іноді виграш однієї сторони означає програш іншої. У цьому випадку доводиться вибирати серед гірших для себе стратегій супротивника найменш гіршу, тобто брати максимум по множині стратегій супротивника і мінімум по власних стратегіях. У цьому полягає мінімаксний критерій, широко використовуваний у теорії ігор. У теорії матричних ігор задають матрицю платежів

$$\left\| a_{ij} \right\|,\ i = \overline{1,m};\ j = 1,n\ .$$

Кожний елемент цієї матриці означає платіж супротивникові у випадку, коли він застосовує стратегію j, а наша сторона - стратегію i. Потрібно знайти серед множини гірших для нас стратегій супротивника найменш погану, тобто вирішити мінімаксну задачу, знайти $\min_i \max_j a_{ij}$.

Неважко переконатися, що якщо змінити знаки a_{ij} на зворотні, а це означає заміну програшу нашої сторони на виграш, то вирішуватиметься максимінна задача знаходження $\max_i \min_j \left(-a_{ij}\right)$. У деяких задачах, що мають так звану **сідлову точку**, $\min_i \max_j a_{ij} = \max_i \min_j a_{ij}$.

Обидві ці задачі оптимізації - мінімаксна і максимінна - не вирішуються шляхом застосування класичних методів.

Задачу управління можна вважати сформульованою математично, якщо сформульовано ціль управління, виражену через критерій ефективності, і визначені обмеження, що є системою алгебраїчних рівнянь або нерівностей, які виражають обмеженість ресурсів або інших величин, використовуваних при управлінні.

Рішення про спосіб управління, що задовольняє всім поставленим обмеженням і перетворює на мінімум (максимум) критерій ефективності, називають **оптимальним рішенням**.

Отже, суть всіх оптимізаційних задач зводиться до пошуку такого розв'язку $\overline{X}$, що перетворює на екстремум критерій ефективності, виражений як функція від елементів прийнятого розв'язку $\overline{X}$ і називаний **цільовою функцією** $F(x)$. Слід зазначити, що математичне програмування є не аналітичною, а алгоритмічною формою розв'язання задач, тобто дає не формулу, яка виражає остаточний результат, а вказує лише обчислювальну процедуру, що призводить до розв'язання задачі. Тому методи математичного програмування стають ефективними головним чином при використанні обчислювальної техніки.

Відзначимо, що до оптимізаційних задач, як правило, незастосовні методи класичного аналізу для відшукання умовних екстремумів. Це пов'язане з такими специфічними їх особливостями:

1) коли на елементи розв'язку $\overline{X}$ накладені обмеження, екстремум часто досягається не в точках, де похідні дорівнюють нулю, а на границі області обмежень;
2) у практичних задачах число змінних і число обмежень настільки велике, що пошук екстремуму шляхом визначення похідних стає не ефективним;
3) у багатьох задачах математичного програмування цільова функція взагалі не має похідних (наприклад, задана тільки для цілочислових значень аргументів).

У зв'язку з цим метою математичного програмування є створення аналітичних методів визначення розв'язку або ефективних обчислювальних способів одержання наближеного розв'язку оптимізаційної задачі.

Математичне моделювання економічних процесів є, з одного боку, дуже важливим і складним, а з іншого боку – таким, що практично не піддається науковій формалізації. Неодноразові спроби, виділити загальні принципи створення математичних моделей призводили або до декларування рекомендацій самого загального характеру, важко застосовних для розв'язання конкретних проблем, або, навпаки, до появи рецептів, застосовних у дійсності тільки до вузького кола задач. Тому кориснішим є знайомство з технікою математичного моделювання на конкретних прикладах.

Як такі приклади розглянемо кілька класичних економіко-математичних моделей і задач, які можуть бути сформульовані на їх основі.

Управління портфелем активів. Розглянемо проблему прийняття інвестором рішення про вкладення капіталу. У таблиці 2.1 показані характеристики об'єктів для інвестування.

Таблиця 2.1 - Характеристики об'єктів інвестування

Найменування	**Прибутко-вість (%)**	**Термін ви-купу (рік)**	**Надійність (бали)**
A	5,5	2011	5
B	6,0	2015	4
C	8,0	2020	2
D	7,5	2012	3
E	5,5	2010	5
F	7,0	2013	4

При ухваленні рішення про придбання активів мають бути дотримані умови:

- сумарний обсяг капіталу, що припускається вкласти, становить 100000 грн;

- частка коштів, що вкладена в один об'єкт, не може перевищувати чверті від усього обсягу;

- понад половину всіх коштів повинні бути вкладені в довгострокові активи (припустимо, що на розглянутий момент до таких належать активи з терміном погашення після 2014 р.);

- частка активів, що мають надійність меншу за 4 бали, не може перевищувати третини від сумарного обсягу.

Складемо математичну модель для даної ситуації. Доцільно почати з визначення структури керованих змінних. У розглянутому прикладі такими змінними є обсяги коштів, вкладені в активи тієї або іншої фірми. Позначимо їх як x_A, x_B, x_C, x_D, x_E, x_F. Тоді сумарний прибуток від розміщених активів, що отримає інвестор, можна записати у вигляді виразу

$$P = 0{,}055x_A + 0{,}06x_B + 0{,}08x_C + 0{,}075x_D + 0{,}055x_E + 0{,}07x_F. \quad (2.1)$$

На наступному етапі моделювання треба формально описати обмеження на структуру портфеля:

- обмеження на сумарний обсяг активів:

$$x_A + x_B + x_C + x_D + x_E + x_F \le 100000. \quad (2.2)$$

- обмеження на розмір частки кожного активу:

$$x_A \le 25000;\ x_B \le 25000;\ x_C \le 25000,$$

$$x_D \le 25000,\ x_E \le 25000,\ x_F \le 25000. \quad (2.3)$$

- обмеження, що пов'язане з необхідністю вкладати половину коштів у довгострокові активи:

$$x_B + x_C \geq 50000. \qquad (2.4)$$

- обмеження на частку ненадійних активів:

$$x_C + x_D \leq 30000. \qquad (2.5)$$

Систему обмежень відповідно до економічного змісту задачі треба доповнити умовами невід'ємності для змінних:

$$x_A \geq 0;\ x_B \geq 0;\ x_C \geq 0;\ x_D \geq 0;\ x_E \geq 0;\ x_F \geq 0. \qquad (2.6)$$

Вирази (2.1)-(2.6) утворюють математичну модель поведінки інвестора. У рамках цієї моделі ставиться задача пошуку таких значень змінних x_A, x_B, x_C, x_D, x_E, x_F, при яких досягається найбільше значення доходу (2.1) і виконуються обмеження на структуру портфеля активів (2.2)-(2.6).

Розглянемо інші загальні моделі і задачі.

Найпростіша задача виробничого планування. Нехай є виробниче підприємство, що виробляє продукцію n видів. У процесі виробництва використовують m видів ресурсів (сировини). Технології виробництва характеризуються нормами витрат сировини на одиницю виробленого продукту. Позначимо через a_{ij} кількість i-го ресурсу ($i = \overline{1-m}$), що витрачається на виробництво одиниці j-го продукту ($j = \overline{1-n}$). Весь набір витрат ресурсів на виробництво j-го продукту можна подати у вигляді вектора-стовпця

$$\overline{A}_j = \begin{bmatrix} a_{1j} \\ a_{2j} \\ \vdots \\ a_{mj} \end{bmatrix},$$

а витрати ресурсів на виробництво всіх n видів продукції - у вигляді прямокутної матриці розмірності m на n:

$$\overline{A} = \begin{bmatrix} a_{11} \ldots a_{1j} \ldots a_{1n} \\ \ldots\ \ldots\ldots\ \ldots\ldots \\ a_{21} \ldots a_{2j} \ldots a_{2n} \\ \ldots\ \ldots\ldots\ \ldots\ldots \\ a_{m1} \ldots a_{mj} \ldots a_{mn} \end{bmatrix}.$$

Якщо j-й продукт виробляється в кількості x_j, то нам потрібно витратити $a_{1j}x_j$ першого ресурсу, $a_{2j}x_j$ - другого, і так далі, $a_{mj}x_j$ - m-го ресурсу. Зведений план виробництва за всіма продуктами можна подати у вигляді n-мірного вектора-рядка $x = (x_1, x_2, ..., x_j, ..., x_n)$. Тоді загальні витрати i-го ресурсу на виробництво всіх продуктів можна виразити як суму

$$\sum_{j=1}^{n} a_{ij} x_j ,$$

що є скалярним добутком векторів A_j, і x. Очевидно, що усяка реальна виробнича система має обмеження на ресурси, які вона витрачає в процесі виробництва.

У рамках розглянутої моделі ці обмеження породжуються m-мірним вектором $b = (b_1, b_2, ..., b_m)$, де b_i - максимальна кількість і-го ресурсу, яку можна витратити у виробничому процесі. У математичній формі ці обмеження представляють у вигляді системи m нерівностей:

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n &\le b_1, \\ a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n &\le b_2, \\ &\ldots\ldots\ldots \\ a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mn}x_n &\le b_m. \end{aligned} \quad (2.7)$$

Застосовуючи правила матричної алгебри, систему (2.7) можна записати в короткій формі, уявивши ліву частину як добуток матриці A на вектор X, а праву - як вектор B:

$$\overline{A}\overline{X} \le \overline{B}\,. \quad (2.8)$$

До системи (2.8) також мають бути додані природні обмеження на невід'ємність компонентів плану виробництва: $x_1 \ge 0$; $x_j \ge 0$; $x_n \ge 0$, або, що те саме,

$$\overline{X} \ge 0\,. \quad (2.9)$$

Позначивши через c_j ціну одиниці j-го продукту, дістанемо вираз сумарного доходу від виконання плану виробництва, що задається вектором $\overline{X}$:

$$\sum_{j=1}^{n} c_j x_j = \overline{C}\overline{X}\,. \quad (2.10)$$

Формули (2.8)-(2.10) є найпростішою математичною моделлю, що описує окремі сторони функціонування певного економічного об'єкта, поведінкою якого ми хочемо управляти. У рамках цієї моделі можна поставити різні задачі, але самою природною буде задача пошуку такого плану виробництва $x \in R^n$, що дає найбільше значення сумарного доходу, тобто функції (2.10), і одночасно задовольняє системі обмежень (2.8)-(2.9).

Незважаючи на явну умовність розглянутої ситуації й простоту задачі, її розв'язання є не тривіальним і багато в чому стало практично можливим тільки після розробки спеціального математичного апарата.

Оскільки будь-яка наукова модель містить спрощуючи передумови, для коректного застосування отриманих за її допомогою результатів необхідно чітке розуміння суті цих спрощень, що дозволяє зробити висновок про їх припустимість або неприпустимість. Найсильнішим спрощенням у розглянутій моделі є припущення про прямо пропорційну (лінійну) залежність між обсягами витрати ресурсів і обсягами виробництва, що задається за допомогою норм витрат a_{ij}. Очевидно, що це припущення далеко не завжди виконується. Так, обсяги витрати багатьох ресурсів (наприклад, основних фондів) змінюються стрибкоподібно залежно від зміни компонентів обсягу виробництва x. До інших передумов, що спрощують, належать припущення про незалежність цін c_j від обсягів x_j, що є справедливим лише для певних меж їх зміни, зневага ефектом кооперації в технологіях та ін. Дані «уразливі» місця важливо знати ще й тому, що вони вказують принципові напрямки вдосконалення моделі.

Транспортна задача. Розглянемо проблему з організації перевезення певного продукту між пунктами його виробництва, кількість яких дорівнює m, і n пунктами споживання. Кожний i-й пункт виробництва ($i=\overline{1,m}$) характеризується запасом продукту $a_i \geq 0$, а кожний j-й пункт споживання ($j=\overline{1,n}$) - потребою в продукті $b_j \geq 0$. Сітку доріг, що з'єднує систему розглянутих пунктів, моделюють за допомогою матриці C розмірності m на n, елементи якої c_{ij} є нормами витрат на перевезення одиниці вантажу з пункту виробництва і до пункту споживання j. План перевезення вантажу подають у вигляді масиву елементів розмірності $m\times n$:

$$x=(x_{11}, \ldots, x_{1n}, x_{21}, \ldots, x_{2n}, \ldots, x_{i1}, \ldots, x_{in}, \ldots, x_{m1}, \ldots, x_{mn}). \qquad (2.11)$$

В (2.11) план перевезень x може розглядатися як вектор, що розпадається на m груп, по n елементів у кожній, причому i-та група відповідає обсягам вантажу, що вивозиться з i-го пункту виробництва до усіх можливих пунктів споживання. Якщо реальне перевезення між пунктами i і j відсутнє, то вважають $x_{ij} = 0$.

Обмеження на можливі значення $x \in R^{mn}$ мають вигляд:

- обмеження на задоволення потреб у всіх пунктах споживання

$$\sum_{i=1}^{m} x_{ij} \geq b_j, \qquad j=\overline{1,n}; \qquad (2.12)$$

- обмеження на можливості вивозу запасів з усіх пунктів виробництва

$$\sum_{j=1}^{n} x_{ij} \leq a_i, \qquad i=\overline{1,m}; \qquad (2.13)$$

- умови невід'ємності компонентів вектора плану

$$x_{ij} \geq 0, \qquad i=\overline{1,m}, \quad j=\overline{1,n}. \qquad (2.14)$$

Істотною характеристикою описуваної моделі є співвідношення параметрів a_i і b_j. Якщо сумарний обсяг виробництва дорівнює сумарному обсягу споживання,

$$\sum_{i=1}^{m} a_i = \sum_{j=1}^{n} b_j ,$$

систему називають **збалансованою**. При цьому розумно накладати такі обмеження на сумарний ввіз і вивіз вантажу, при яких повністю вивозиться весь вантаж і не залишається незадоволених потреб, тобто умови (2.12) і (2.13) виконуються як рівності.

За аналогією із задачею виробничого планування припустимо, що витрати на перевезення прямо пропорційні кількості перевезеного вантажу. Тоді сумарні витрати на перевезення в системі приймуть вигляд:

$$f(x) = \sum_{i=1}^{m} \sum_{j=1}^{n} c_{ij} x_{ij} . \qquad (2.15)$$

Функція (2.15) і описані вище обмеження задають транспортну модель. На її основі можна сформулювати задачу мінімізації сумарних витрат на перевезення, яка у літературі одержала назву транспортної задачі в матричній постановці. Загалом кажучи, транспортна задача є окремим випадком найпростішої задачі виробничого планування, але в силу ряду особливостей для її розв'язання застосовують специфічні методи, які, крім того, дозволяють прийти до важливих теоретичних узагальнень.

Спільним для розглянутих вище задач є те, що в них стоїть проблема з пошуку найбільшого або найменшого (**оптимального**) значення певної функції, що відбиває **ціль управління** системою, або, як ще кажуть, **цільової функції**. Пошук оптимального значення здійснюється на певній підмножині припустимих значень змінних, що описують стан цієї системи, іменованій **множиною припустимих планів**.

Нехай на певній множині D визначена функцію $f(x)$. Нагадаємо, що точку x^*, що належить D ($x^* \in D$), називають **точкою глобального максимуму**, якщо для будь-якого $x \in D$ виконується нерівність $f(x) \le f(x^*)$. У цьому випадку значення $f(x^*)$ називають **глобальним максимумом функції**. Точку $\hat{x}$ називають **точкою локального максимуму**, якщо існує певне оточення цієї точки, у будь-якій точці якого значення функції менші, ніж в $\hat{x}$ ($f(x) \le f(\hat{x})$). Аналогічно визначають **глобальний** і **локальний мінімуми**. Узагальнюючим поняттям для максимуму й мінімуму є такий термін, як **екстремум (оптимум)**.

Необхідно відзначити, що далеко не завжди весь комплекс цілей і задач, що стоїть перед об'єктом, який моделюють, можна виразити у формі певної цільової функції. Усвідомлення цієї проблеми стало переломним етапом в історії розвитку науки і дало поштовх до розвитку нових напрямків, пов'язаних з методами **багатокритеріальної (або векторної) оптимізації**, коли критерієм оптимальності є вимога мінімізації або максимізації кількох скалярних функцій. Проте всі вони базуються на методах **однокритеріальної оптимізації**, без ясного розуміння яких неможлива робота із складнішим математичним апаратом.

Потужним інструментом вирішення подібних задач стали спеціальні методи пошуку екстремуму, що складають зміст математичного програмування. У цьому випадку поняття «програмування» вживається в значенні «планування» (на відміну від програмування для ЕОМ).

Оптимізаційні задачі мають велику розмаїтість. Математичне моделювання цих задач практично не піддається науковій формалізації через те, що принцип побудови математичної моделі істотно залежить від конкретної природи досліджуваної системи. Проте у цих задачах прийнято виділяти певну послідовність етапів дослідження:

1. Постановка задачі.
2. Словесне формулювання задачі з визначенням цілі її розв'язання і факторів-обмежень, що впливають на нього (вербальна модель);
3. Формалізація задачі – побудова адекватної математичної моделі. На цьому етапі цільову функцію $F(x)$ виражають як залежність від розв'язку $\overline{X}$, а обмеження записують у вигляді системи рівностей і нерівностей;
4. Розв'язання задачі на базі математичної моделі;
5. Перевірка отриманих результатів на їх адекватність природі досліджуваної системи, можливе коректування первісної моделі.

6. Розробка рекомендацій на підставі отриманого розв'язку.

Введемо ряд визначень.

Розв'язком (або **планом**) називають певний вибір залежних від нас параметрів. Параметри, сукупність яких утворює розв'язок, називають **елементами розв'язку**. Як елементи розв'язку можуть фігурувати числа, вектори, функції, фізичні ознаки та ін.

$$x = (x_1, x_2, \dots, \dots, x_j, \dots, x_n).$$

Система обмежень за ресурсами формує множину припустимих розв'язків (планів) D. Той факт, що розв'язок x належить множині припустимих розв'язків D, записується в такий спосіб:

$$x \in D.$$

Оптимальним розв'язком або **оптимальним планом** називають таке рішення, що обертає цільову функцію $F(x)$ на максимум або мінімум.

Отже, у найбільш загальному вигляді оптимізаційна задача формулюється в такий спосіб:

> При заданих обмеженнях знайти такий розв'язок $x = x^*$, що обертає цільову функцію $F(x)$ на максимум або мінімум.
>
> $$F^* = \underset{x \in D}{extr}\{F(x, \alpha)\},$$
>
> де α - система обмежень задачі.

Залежно від вигляду цільової функції $F(x)$ і системи обмежень α у математичному програмуванні виділяють наступні методи.

Лінійне програмування. Застосовують, якщо в моделі цільова функція $F(x)$ є лінійною, а множина D, на якій шукають її екстремум, заданий системою лінійних рівностей і нерівностей. У лінійному програмуванні існує клас задач, структура яких дозволяє створити спеціальні методи розв'язання, що вигідно відрізняються від методів розв'язання задач загального характеру. Зокрема, транспортна задача.

Нелінійне програмування. Тут є нелінійними цільова функція й обмеження. У нелінійному програмуванні виділяють такі класи задач:

- опукле програмування – коли цільова функція є опуклою (якщо розглядають задачу її мінімізації) і опуклою є множина, на якій вирішується екстремальна задача;

- квадратичне програмування – коли цільова функція є квадратичною, а обмеження – лінійні рівності або нерівності.

Дискретне програмування. Цей метод використовують, коли на елементи рішення x накладено вимогу дискретності, наприклад, цілочисленості. Така вимога істотно ускладнює розв'язання задачі, тому що застосування стандартних прийомів (вирішити задачу як аналогову, а потім округлити результат до цілого значення) неможливо.

Динамічне програмування. Це метод, що дозволяє шляхом покрокової оптимізації певних проміжних цільових функцій отримати загальний результуючий оптимум. У задачах динамічного програмування цільова функція $F(x)$ є адитивною або мультиплікативною функцією змінних x.

Стохастичне програмування. Даний вид програмування використовують, коли параметри умов або елементи розв'язку є випадковими величинами, що зумовлене невизначеністю, яка породжує ризикованість прийнятих рішень. У стохастичному програмуванні труднощі виникають не тільки при розробці методів розв'язання задач, а й при їх постановці.

Евристичне програмування. Застосовують для розв'язання задач, у яких точний оптимум знайти алгоритмічним шляхом неможливо через величезне число варіантів. У такому випадку відшукують не оптимальний, а досить гарний з погляду практики розв'язок.

Контрольні запитання

1. Охарактеризуйте особливості оптимізаційних задач.
2. Які загальні етапи розв'язання оптимізаційних задач прийнято виділяти?
3. Чому до оптимізаційних задач не застосовують класичні методи пошуку умовного екстремуму функції?
4. Що являє собою цільова функція оптимізаційної задачі? Яке її призначення?
5. Дайте визначення понять: план, припустимий план, оптимальний план, розв'язок оптимізаційної задачі.
6. На чому заснована класифікація моделей і методів математичного програмування з розв'язання оптимізаційних задач? Які класи моделей і методів виділяють у математичному програмуванні?
7. Поясніть, що є множиною можливих розв'язків задачі математичного програмування?
8. Поясніть, яку область можливих розв'язків задачі математичного програмування називають областю припустимих планів.

ТЕМА 3
ЛІНІЙНЕ ПРОГРАМУВАННЯ

Серед задач математичного програмування найпростішими й найкраще розробленими є задачі лінійного програмування. Характерним для них є те, що:

- цільова функція $F(x)$ лінійно залежить від елементів розв'язку $x_1, x_2, \ldots, x_n$. Її називають **лінійною формою** й позначають L;
- обмеження, що накладають на елементи розв'язку, мають вигляд лінійних рівностей і нерівностей відносно $x_1, x_2, \ldots, x_j, \ldots, x_n$.

Такі задачі часто зустрічаються на практиці. До задач лінійного програмування належить, зокрема, розглянута раніше найпростіша задача виробничого планування (задача про оптимальне використання ресурсів). Також до задач лінійного програмування належать задачі про використання інвестицій, про мінімізацію витрат, транспортна задача, задача про складання раціону та ін.

Математичну модель задачі лінійного програмування завжди записують у двох формах – у **загальній формі** (ЗЗЛП) і **канонічній формі** (КЗЛП).

3.1 Загальна форма задачі лінійного програмування (ЗЗЛП)

У загальному вигляді задачу лінійного програмування (ЗЛП) формулюють в такий спосіб:

Знайти найбільше або найменше значення лінійної функції

$$L = \sum_{j=1}^{n} c_j x_j \to extr \,. \tag{3.1}$$

на певній множині D, де $x \in D$ задовольняє системі обмежень

$$\begin{gathered}
a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n \le b_1, \\
a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n \le b_2, \\
\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\
a_{k1}x_1 + a_{k2}x_2 + \ldots + a_{kn}x_n \le b_k, \\
a_{(k+1)1}x_1 + a_{(k+1)2}x_2 + \ldots + a_{(k+1)n}x_n = b_{(k+1)}, \\
\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\
a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mn}x_n = b_m, \\
x_j \ge 0, \quad j = \overline{1, n} \,.
\end{gathered} \tag{2.2}$$

Відзначимо, що в системі (2.2) перші k обмежень є нерівностями, а наступні m-k - рівняннями. Цього можна домогтися простим переупорядкуванням виразів.

Щодо направлення знака нерівності вважатимемо, що ліва частина менша або дорівнює правій частині. Домогтися цього можна, помноживши на (-1) обидві частини нерівностей із протилежним знаком.

Вибір типу шуканого екстремуму цільової функції також не принциповий, оскільки задача пошуку максимуму функції $L = \Sigma c_j x_j$ є еквівалентною задачі пошуку мінімуму функції $-L = \Sigma(-c_j x_j)$.

Задачу лінійного програмування, записану в такій формі, називають загальною задачею лінійного програмування (ЗЗЛП).

3.2 Основні властивості ЗЗЛП та її перша геометрична інтерпретація

У тому випадку, коли ЗЗЛП містить дві змінні x_1 і x_2, її можна зобразити на координатній площині й одержати розв'язок графічним методом. Графічне розв'язання ЗЗЛП носить ілюстративний характер, але основний зміст і термінологія розповсюджуються на задачі великої розмірності.

Розглянемо приклад. Нехай цільова функція представлена виразом

$$L = x_1 + 3x_2 \to \max,$$

а обмеження задані системою нерівностей:

$$\begin{gathered}
x_1 + x_2 \le 6, \\
x_1 - x_2 \le 2, \\
x_1 \qquad \le 3, \\
x_1 \ge 0, x_2 \ge 0.
\end{gathered}$$

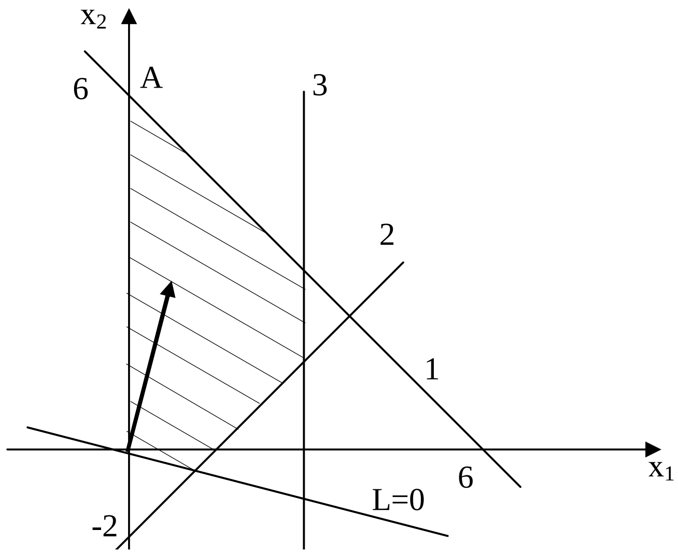


Рисунок 3.1 – Геометрична інтерпретація ЗЗЛП

Будемо зображувати пару значень x_1 і x_2 точкою на координатній площині $x_1 0 x_2$ з координатами (x_1, x_2), що показано на рисунку 3.1.

Кожна нерівність визначає певну напівплощину. Перетинання трьох напівплощин є множиною припустимих планів D, тому що кожна точка його множини належить одночасно кожній з трьох напівплощин, а отже задовольняє обмеженням ЗЗЛП. Помітимо, що припустимих розв'язків - нескінченна кількість.

Для визначення оптимального плану задачі, тобто такого розв'язку (x_1, x_2), що обертає цільову функцію на максимум, скористаємося визначеннями:

- градієнтом функції $f(x)$ називають вектор

$$\nabla f(x) = \left(\frac{\partial f(x)}{\partial x_1}, \dots, \frac{\partial f(x)}{\partial x_n} \right),$$

який вказує напрям найскорішого зростання функції $f(x)$;

- лінією рівня функції $f(x)$ називають множину точок з області її визначення, у яких функція приймає те саме фіксоване значення.

У нашому прикладі $\nabla L = (1, 3)$. Лінії рівня L перпендикулярні напрямку градієнта. Побудуємо опорну пряму L=0, що проходить через початок координат, і будемо переміщувати її в напрямку ∇L. Очевидно, що переміщувати лінію рівня в напрямку зростання цільової функції є смисл тільки в межах області припустимих розв'язків. Точкою, у якій цільова функція дістане максимального значення, у нашому прикладі є точка A з координатами (0, 6). Отже, отриманий оптимальний план задачі

$$x^* = (0, 6),$$

за якого цільова функція приймає максимальне значення

$$L = 1*0 + 3*6 = 18.$$

Теоретично можливі також такі окремі випадки розв'язку ЗЗЛП:

- цільова функція L не обмежена зверху, тобто не має максимуму (рис. 3.2);

- коли лінія рівня збігається з гранню області припустимих розв'язків (рис. 3.3). У цьому випадку всі точки, що лежать на грані множини D, є оптимальними планами й кажуть, що має місце **альтернативний оптимум**.

У розглянутих ілюстраціях припустимі плани ЗЗЛП мають вигляд опуклої багатогранної множини. Таке подання множини припустимих планів називають першою геометричною інтерпретацією ЗЗЛП.

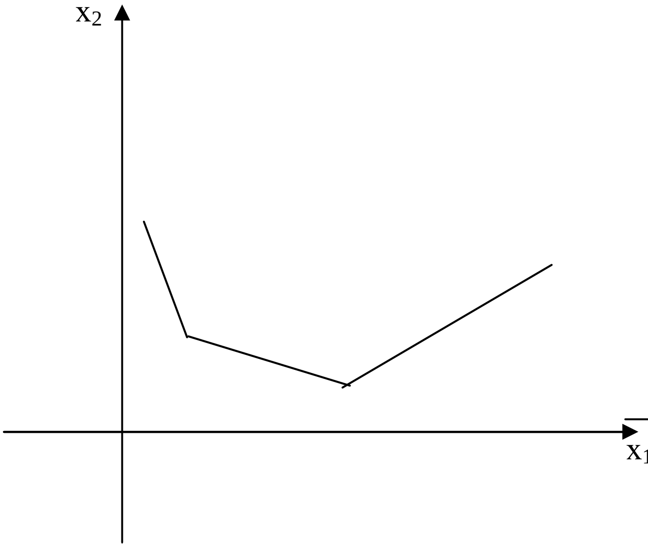


Рисунок 3.2 - Цільова функція *L* не обмежена зверху

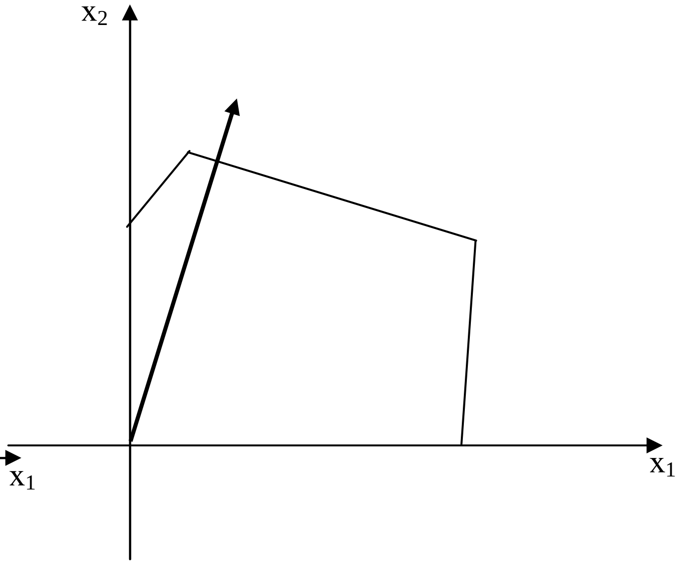


Рисунок 3.3 – Альтернативний оптимум

3.3 Основні теореми лінійного програмування

Розглянемо деякі теореми, що відбивають фундаментальні властивості задач лінійного програмування і полягають в основі методів їх розв'язання. Вони узагальнюють на випадок задач із довільною кількістю змінних ті властивості, які ми спостерігали у двовимірному випадку.

> **Теорема 3.1.** Якщо цільова функція *L* приймає максимальне значення в певній точці множини припустимих планів *D*, то вона приймає це значення й у певній кутовій точці цієї множини.

Доказ.

Щоб не ускладнювати виклад, обмежимося тим випадком, коли множина *D* є обмеженою, і, отже, є опуклим багатогранником. Для доказу скористаємося наступною відомою властивістю обмежених опуклих множин:

> якщо *D* - замкнена обмежена опукла множина, що має кінцеве число кутових точок, то будь-яку точку $x \in D$ можна подати у вигляді опуклої комбінації кутових точок *D*.

Нехай $x_1, x_2, \dots, x_m$ - кутові точки множини *D*, а x^* - точка, у якій цільова функція *L* досягає максимуму. Точку x^* можна подати у вигляді опуклої комбінації кутових точок $x_1, x_2, \dots, x_m$

$$x^* = \sum_{i=1}^{m} \lambda_i x_i ,$$

де $\sum \lambda_i = 1, \quad \lambda_i \geq 0, \quad i = \overline{1, m}$.

Оскільки x^* - точка максимуму, то для будь-якого x $cx^* \geq cx$. У тому числі й для cx_r (x_r – кутова точка).

Функція $L(x)$ - лінійна, тому

$$L\left(\sum_{i=1}^{m} \lambda_i x_i\right) = \sum_{i=1}^{m} \lambda_i L(x_i),$$

а тоді

$$cx^* = c\sum_{i=1}^{m} \lambda_i x_i = \sum \lambda_i (cx_i) \leq \sum \lambda_i (cx_r) = cx_r \sum \lambda_i = cx_r,$$

де x_r – кутова точка множини D, що задовольняє умові

$$cx_r = \max_{1<i<m} \{cx_i\}.$$

Отже, $cx^* \leq cx_r$. У той самий час $cx^* \geq cx_r$, звідки випливає $cx^* = cx_r$.

Тобто існує принаймні одна кутова точка x_r, у якій цільова функція приймає максимальне значення.

Теорема 3.2. Якщо цільова функція L приймає максимальне значення в кількох точках множини D, то вона приймає це саме значення в будь-якій точці, що є їх опуклою комбінацією.

Доказ.

Нехай максимальне значення цільової функції $L(x)$ досягається в точках $x_1, x_2, \ldots, x_s$, тобто $L^* = cx_i, \quad i = \overline{1, s}$. Розглянемо довільну опуклу комбінацію цих точок

$$x^* = \sum_{i=1}^{s} \lambda_i x_i,$$

де $\sum \lambda_i = 1, \quad \lambda_i \geq 0, \quad i = \overline{1, s}$.

Знайдемо значення цільової функції в точці x^*

$$L(x^*) = cx^* = c\sum_{i=1}^{s} \lambda_i x_i = \sum \lambda_i (cx_i) = \sum \lambda_i L^* = L^* \sum \lambda_i = L^*.$$

Теорема 3.1 є фундаментальною, тому що вона вказує принциповий шлях розв'язання ЗЛП. Замість дослідження нескінченної множини припустимих розв'язків для знаходження серед них оптимального, необхідно досліджувати лише кінцеве число кутових точок багатогранника розв'язків.

3.4 Канонічна форма задачі лінійного програмування (КЗЛП)

Канонічною називають задачу лінійного програмування, якщо всі її обмеження є рівняннями.

Переважна більшість методів розв'язання задач лінійного програмування призначена для канонічних задач. Тому початковий етап розв'язання будь-якої загальної ЗЛП завжди пов'язаний із приведенням її до еквівалентної КЗЛП.

Загальна ідея переходу від ЗЗЛП до КЗЛП досить проста. Обмеження у вигляді нерівностей перетворюють на рівняння за рахунок додавання фіктивних невід’ємних змінних, які одночасно входять до цільової функції з коефіцієнтом 0, тобто не надають впливу на її значення.

$$x'_{n+i}, \quad i = \overline{1,k}.$$

Змінні, на які не накладено умову невід’ємності, подаються у вигляді різниці двох нових невід’ємних змінних.

Додатково треба помітити, що вибір типу шуканого екстремуму (максимуму або мінімуму) носить відносний характер. Так, задача пошуку максимуму функції

$$L(x) = \sum_{j=1}^{n} c_j x_j$$

є еквівалентною задачі пошуку мінімуму функції

$$-L(x) = \sum_{j=1}^{n} -c_j x_j .$$

У результаті таких перетворень одержують канонічну задачу:

Знайти найбільше або найменше значення функції

$$L = c_1x_1 + c_2x_2 + \ldots + c_n x_n + 0 \cdot x'_{n+1} + \ldots + 0 \cdot x'_{n+k} \quad (3.3)$$

на певній множині D, де $x \in D$ задовольняє системі обмежень

$$\begin{aligned}
a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n + x'_{n+1} &= b_1,\\
a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n + x'_{n+2} &= b_2,\\
\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots&\ldots\ldots\\
a_{k1}x_1 + a_{k2}x_2 + \ldots + a_{kn}x_n + x'_{n+k} &= b_k,\\
a_{k+1,1}x_1 + a_{k+1,2}x_2 + \ldots + a_{k+1,n}x_n &= b_{k+1},\\
\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots&\ldots\ldots\\
a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mn}x_n &= b_m,\\
x_j \ge 0, \quad j = \overline{1, n+k}&.
\end{aligned} \quad (3.4)$$

Можна показати, що отримана задача еквівалентна вихідній задачі.

Отже, «платою» за перехід від загальної форми задачі лінійного програмування до канонічної є зростання її розмірності, що, за інших рівних умов, ускладнює процес розв'язання.

3.5 Базисні розв'язки канонічної задачі лінійного програмування

Запишемо КЗЛП у матричній формі:

знайти максимум

$$\overline{C}^T \overline{X} \quad (3.5)$$

за умови

$$\overline{A}^T \overline{X} = \overline{B}, \qquad \overline{X} \ge 0, \quad (3.6)$$

де $\overline{A}$, $\overline{X}$ і $\overline{B}$ – вектори.

$$\overline{A_1} = \begin{pmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{m1} \end{pmatrix}; \qquad \overline{A_2} = \begin{pmatrix} a_{12} \\ a_{22} \\ \vdots \\ a_{m2} \end{pmatrix} \qquad \overline{A_n} = \begin{pmatrix} a_{1n} \\ a_{2n} \\ \vdots \\ a_{mn} \end{pmatrix}; \qquad \overline{B} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix}. \qquad (3.7)$$

Будемо розглядати стовпці матриці $\overline{A}_j$ як вектори простору R^m. Тоді кожному припустимому плану КЗЛП (n-мірному вектору $\overline{X}$) відповідає лінійна комбінація стовпців $\overline{A}_j$, що дорівнює стовпцю $\overline{B} \in R^m$.

Вектори $\overline{A}_j$, $j = \overline{1,n}$ називають векторами вимог задачі. Вектор $\overline{B}$ називають вектором обмежень. При цьому компоненти певного припустимого плану $x \in D$ – x_j – є коефіцієнтами розкладання вектора обмежень $\overline{B}$ задачі по векторах вимог $\overline{A}_j$.

Очевидно, що кількість рівнянь, які задають множину D, менша або дорівнює кількості змінних задачі ($m \le n$). Якщо це не так, то або система рівнянь

$$\overline{A}\ \overline{X} = \overline{B}$$

є несумісною, або вона містить надлишкові (лінійно залежні) рівняння.

Якщо певні m стовпців A_1, A_2, …, A_m матриці $\overline{A}$ є лінійно незалежними, то вони утворюють базис у просторі R^m і їх буде достатньо для подання вектора $\overline{B}$ у вигляді лінійної комбінації зазначених стовпців. Це означає, що інші стовпці ввійдуть до даного розкладання з нульовими коефіцієнтами.

Якщо коефіцієнти лінійної комбінації опиняться невід'ємними, то ми отримаємо **базисний припустимий план** x, в якого не більш за m елементів відмінні від нуля. Або іншими словами. Система обмежень КЗЛП є системою m рівнянь з n змінними, причому $m \le n$.

У такій системі m змінних називають **базисними**, а інші (n-m) змінних – **вільними**.

Базисним розв'язком системи m лінійних рівнянь з n змінними називають розв'язок, у якому всі n-m вільних змінних дорівнюють нулю.

Базисними розв'язками можуть бути різні групи з n змінних. У загальному випадку число таких груп не перевершує C_n^m.

Отже, система з m лінійних рівнянь з n змінними ($m<n$) є невизначеною, тому що кожному довільному набору вільних змінних відповідає один базисний розв'язок системи.

Опорним планом КЗЛП називають припустимий базисний розв'язок, компоненти якого перевищують нуль.

Базисний розв'язок, в якому хоча б одна з базисних змінних дорівнює нулю, називають **виродженим**.

Властивості базисних планів задачі лінійного програмування виражає наступна теорема

Теорема 3.3. Кожний припустимий базисний план є кутовою точкою множини припустимих планів *D*.

Доказ.

Вважатимемо, що базисними є перші *m* стовпців матриці $\overline{A}$

$A_1, A_2, \ldots, A_m$...

Тоді формулювання теореми можна записати у вигляді:

Якщо існує такий *n*-мірний вектор

$$x = \left(x_1, x_2, \ldots, x_k, \underbrace{0, \ldots, 0}_{n-k} \right),$$

$$x_j > 0; \quad j = \overline{1,k}; \quad k \le m,$$

що $x_1 A_1 + x_2 A_2 + \ldots + x_k A_k = B_0$, то *x* є кутовою точкою множини *D*.

Припустимо, що базисний план *x* не є кутовою точкою множини *D*. У цьому випадку його можна подати у вигляді опуклої комбінації двох різних припустимих планів x^1 і x^2

$$x = \lambda x^1 + (1-\lambda)x^2, \quad 0 < \lambda < 1.$$

Оскільки останні (*n*-*k*) компонент вектора *x* дорівнюють нулю, а λ і (1-λ) перевищують нуль, то ці самі (*n*-*k*) компонент у векторах x^1 і x^2 також дорівнюють нулю.

Оскільки плани x^1 і x^2 – припустимі плани, маємо

$$x_1^1 A_1 + x_2^1 A_2 + \ldots + x_k^1 A_k = B,$$

$$x_1^2 A_1 + x_2^2 A_2 + \ldots + x_k^2 A_k = B.$$

Віднімемо з першого рівняння друге

$$(x_1^1 - x_1^2)A_1 + (x_2^1 - x_2^2)A_2 + \ldots + (x_k^1 - x_k^2)A_k = 0.$$

Отримали нульовий вектор. Оскільки A_j лінійно незалежні, нулю дорівнюють коефіцієнти

$$(x_1^1 - x_1^2) = 0,$$

$$(x_2^1 - x_2^2) = 0,$$

$$\vdots$$

$$(x_k^1 - x_k^2) = 0.$$

Звідки випливає, що $x^1 = x^2$. Це суперечить припущенню, що x^1 і x^2 є різними кутовими точками множини *D*.

Отже, план *x* не може бути поданим у вигляді опуклої комбінації двох інших точок множини, і є кутовою точкою даної множини.

Справедливо й зворотне твердження: Якщо *x* - кутова точка множини *D*, то вона є припустимим базисним планом задачі лінійного програмування.

3.6 Симплекс-метод

На підставі розглянутих теорем про властивості лінійних екстремальних задач можна побачити, що пошук їх розв'язків зводиться до послідовного перебору кутових точок множини припустимих планів.

Однак такий перебір для реальних багатомірних задач на практиці нездійсненний або вкрай неефективний. Наприклад, число базисних планів у задачі з 10 змінними й 30 обмеженнями становить

$$C_{30}^{10} = \frac{30!}{10!*20!} = 1489411.$$

Класичним методом розв'язання задач лінійного програмування є симплекс-метод, що також називають методом послідовного поліпшення плану. Він розроблений в 1947 році американським математиком Джорджем Данцигом.

Симплекс-метод є послідовним перебором кутових точок області припустимих розв'язків, при якому значення цільової функції зростає від ітерації до ітерації (від однієї кутової точки до іншої).

Критерій оптимальності в симплекс-методі реалізується шляхом визначення спеціальних оцінок Δ_j для небазисних векторів-стовпців матриці $\overline{A}$, щодо поточного базису (симплекс-різниць). Симплекс-різниці обчислюють за формулою

$$\Delta_j = L_j - c_j, \tag{3.8}$$

де L_j – індекси векторів, що відповідають поточному базису

$$L_j = \sum_{i=1}^{m} c_j a_{ij} . \tag{3.9}$$

Сформулюємо критерій оптимальності припустимого базисного плану:

план є оптимальним, якщо для всіх $j = \overline{1,n}$ $\Delta_j \geq 0$, *і неоптимальним у протилежному випадку, тобто якщо існує таке* $j = \overline{1,n}$, *що* $\Delta_j < 0$.

Якщо симплекс-різниці показують неоптимальність плану, здійснюється перехід до наступного базису. При цьому один стовпець виводиться з базису, а інший вводиться. Для забезпечення покращення значення цільової функції в базис повинен бути введеним вектор-стовпець, що має від'ємну оцінку. Якщо таких стовпців кілька, то для введення рекомендується вибирати стовпець, який має максимальний за модулем добуток оцінки Δ_j на відношення $\Theta_r = \min\limits_i \left\{ \frac{b_i}{a_{ij}} \right\}$. Одночасно на цьому кроці потрібно ухвалити рішення щодо того, який стовпець треба вивести з базису. Зробити це потрібно так, щоб знову сформований базис опинився припустимим.

Можна довести, що припустимість базису, до якого здійснюється перехід, забезпечується наступним правилом виводу стовпця з поточного базису:

для стовпця, що претендує на введення до базису, і вектора обмежень $\overline{B}$ розглядають відношення

$$\Theta_i = \frac{b_i}{a_{ij}}. \tag{3.10}$$

і визначають такий рядок r, що

$$\Theta_r = \min_i \{\Theta_i\}. \tag{3.11}$$

Отриманий індекс r визначає номер рядка, який відповідає вектору, виведеному з базису.

Отже, якщо базис на q-ї ітерації включав стовпці з номерами

$$\{j_1, j_2, ..., j_{r-1}, j_r, j_{r+1}, ..., j_m\},$$

то базис на ітерації $q + 1$ складатиметься зі стовпців з номерами:

$$\{j_1, j_2, ..., j_{r-1}, j_{r+1}, ..., j_m\}.$$

Окремо слід домовитися про випадок, коли стовпець, що претендує на введення до базису, не містить додатних компонентів (a_{ij}<0). Це означає, що цільова функція у задачі не обмежена на множині припустимих значень, тобто може досягати як завгодно великого значення, отже оптимальний план відсутній.

Після переходу до нового базису можна заново сформувати матрицю $\overline{A}$ й дослідити новий план на оптимальність. З погляду техніки обчислень раціонально безпосередньо переходити від матриці q-ї ітерації до матриці (q+1)-ї ітерації. Справа в тому, що в цих матриць стовпці, які відповідають базисним векторам, складаються з нулів, за винятком одного елемента, що дорівнює одиниці. Позиція цього ненульового (одиничного) елемента визначається порядковим номером базисного стовпця. Тому для одержання матриці (q+1)-ї ітерації досить за допомогою лінійних операцій над рядками матриці q-ї (попередньої) ітерації привести її стовпець, що відповідає вектору, який вводиться до базису, до «базисного» вигляду.

Для цього застосовують перетворення Жордана-Гауса (так званий метод повного виключення). У цьому випадку воно полягає в тому, що ми повинні отримати одиницю на місці елемента a_{rj} (його називають ведучим) і нулі на місці інших елементів стовпця a_{ij}. Перше досягається за допомогою поділення r-го рядка на ведучий елемент, друге - шляхом додавання знову отриманого r-го рядка, помноженого на відповідний коефіцієнт, до інших рядків матриці q-ї (попередньої) ітерації.

Формально результат виконання даного перетворення над елементами матриці може бути виражений у наступному виді:

$$a_{rj}^{q+1} = \frac{a_{rj}^q}{a_{rk}^q}, \qquad b_r^{q+1} = \frac{b_r^q}{a_{rk}^q}, \tag{3.12}$$

де $j = \overline{1, n}$;

k – номер стовпця, що вводиться до базису;

$$a_{ij}^{q+1} = a_{ij}^q - a_{ik}^q \frac{a_{rj}^q}{a_{rk}^q}, \tag{3.13}$$

де $i = \overline{1, m}$, $i \neq r$;

$$b_i^{q+1} = b_i^q - a_{ik}^q \frac{b_r^q}{a_{rk}^q}, \tag{3.14}$$

де $i = \overline{1,m}$, $i \neq r$.

Слід особливо відзначити зміст елементів вектора $\overline{B}$. Його нульовий компонент b_0 містить значення цільової функції, якого вона досягає на поточному плані, а інші елементи є ненульовими компонентами цього плану.

Наведемо схему алгоритму симплекс-методу для розв'язання задачі максимізації.

1. Знаходять припустимий базисний план.

2. Перевіряють оптимальність поточного базисного плану: здійснюють перегляд рядка оцінок Δ_j. Можливі два варіанти:

- $\Delta_j \geq 0$ - план, що відповідає поточному базису задачі, є оптимальним. Обчислювальний процес закінчують. Виписують оптимальний план задачі x^* і значення цільової функції $L(x^*)$;

- у рядку оцінок Δ_j існує щонайменше один елемент $\Delta_j < 0$, тобто оцінка якого є від'ємною. Отже, план є неоптимальним. Вибирають стовпець із номером k, для якого добуток $\Delta_j\ \Theta$ є найбільшим за абсолютною величиною. Він називається ведучим і повинен бути введений до чергового базису;

3. Визначають стовпець, що треба вивести з базису. Досліджують ведучий стовпець, можливі два варіанти:

- для всіх $i = \overline{1,m}$ a_{ik}^q < 0. Роблять висновок про необмеженість цільової функції й завершують обчислювальний процес.

- існує принаймні один рядок з номером $i = \overline{1,m}$, для якого a_{ik}^q >0. Відповідно до правила (3.11) визначають номер r стовпця, який буде виведеним з базису;

4. Перераховують елементи матриці $\overline{A}$ й стовпця $\overline{B}$ щодо нового базису відповідно до формул (3.12)-(3.14). Переходять до пункту 2 алгоритму.

Таблична реалізація симплекс-методу. З погляду забезпечення раціональності і наочності обчислювального процесу виконання алгоритму симплекс-методу зручно оформляти у вигляді послідовності таблиць. У різних джерелах наводяться різні модифікації симплекс-таблиць, що відрізняються одна від одної розташуванням окремих елементів. Однак всі вони базуються на тих самих принципах. Зупинимося на структурі таблиці, що наведена у таблиці 3.1.

Таблиця 3.1 - Структура симплекс-таблиці

Базис	$C_{баз}$	C_j					
		B	A_1	A_2	A_j	…	A_n
A_j							
	L_j Δ_j						

Стовпець «Базис» містить найменування базисних векторів (у тій послідовності, у якій вони входять до базису), стовпець $C_{баз}$ – містить коефіцієнти цільової функції при базисних змінних, стовпець B - компоненти вектора обмежень щодо поточного базису, A_1-A_n - компоненти матриці задачі щодо поточного базису. У рядку L_j записують індекси, визначені за формулою (3.9).

У рядку Δ_j містяться поточні оцінки стовпців. Рядок C_j містить коефіцієнти при компонентах поточного плану в цільовій функції.

Слід зазначити, що таблична модифікація симплекс-методу має важливе практичне значення не стільки як зручна форма організації ручного розрахунку, скільки як основа для реалізації даного алгоритму в рамках програмного забезпечення ЕОМ.

Розглянемо приклад розв'язання ЗЛП симплекс-методом. Нехай дана канонічна задача ЛП:

$$L(x)= 50x_1 - 10x_2 + 6x_3 + 40x_4 - 30x_5 \to \max,$$
$$x_1 +x_4+ 3x_5 = 12,$$
$$2x_1 + x_2 +3x_4 = 14,$$
$$-2x_1 + 3x_3 - 4x_4 = 17,$$
$$x_1>0,\ x_2>0,\ x_3>0,\ x_4>0,\ x_5>0.$$

Як видно, стовпці матриці з номерами {5, 2, 3} є лінійно незалежними. І можна одержати розкладання по даних стовпцях вектора обмежень із додатними коефіцієнтами. Останнє означає, що стовпці {5, 2, 3} утворюють припустимий базис, з якого можна почати розв'язання задачі. Початковий опорний план має вигляд x^1 = {0, 14, 17/3, 0, 4}. Заповнюємо симплекс-таблицю 3.2, що відповідає першої ітерації (q=1).

Таблиця 3.2 - Перша ітерація (q=1)

Базис	$C_{баз}$	C_j	50	-10	6	40	-30
		B	A_1	A_2	A_3	A_4	A_5
A_2	-10	14	2	1	0	3	0
A_3	6	17/3	-2/3	0	1	-4/3	0
A_5	-30	4	1/3	0	0	1/3	1
	L_j	-226	-34	-10	6	-48	-30
	Δ_j		-84	0	0	-88	0

Оскільки рядок оцінок у першому й четвертому стовпцях містить від’ємні елементи Δ_1 = -84, Δ_4= -88, план x^1 = {0, 14, 17/3, 0, 4} не є оптимальним, і значення цільової функції $L(x^1)$ = -226 можна покращити. Перейдемо до нового опорного плану.

Визначимо вектор, що будемо вводити до базису (A_1 або A_4).

Відповідно до алгоритму симплекс-методу визначимо відношення Θ. Для вектора A_1: $\Theta_r = \min\limits_i \left\{\frac{14}{2}; \frac{4}{1/3}\right\} = \min\{7;12\} = 7$, r=2; для вектора A_4: $\Theta_r = \min\limits_i \left\{\frac{14}{3}; \frac{4}{1/3}\right\} = \min\{4,66;12\} = 4,66$, r=2. Добутки для вектора A_1:

-34*7= -238; для вектора A_4: -48*4,66= -223,7. Вважаємо номер стовпця, що вводиться до чергового базису, $k = 1$ (тому що |-238|>|-223,7|). З базису повинен бути виведений стовпець з номером 2. Отримаємо черговий припустимий базис задачі {1, 3, 5}. Елемент, що перебуває на перетинанні виділених стовпця й рядка таблиці є ведучим, він дорівнює 2. Застосувавши формули (3.12)-(3.14), переходимо до симплекс-таблиці 3.3, що відповідає другої ітерації, і вважаємо індекс поточної ітерації $q = 2$.

Таблиця 3.3 - Друга ітерація

Базис	$C_{баз}$	C_j	50	-10	6	40	-30
		B	A_1	A_2	A_3	A_4	A_5
A_1	50	7	1	1/2	0	3/2	0
A_3	6	10,3	0	0,2	1	-0,3	0
A_5	-30	1,7	0	-0,2	0	-0,2	1
	L_j	360,8	50	32,2	6	79,2	-30
	Δ_j		0	42,2	0	39,2	0

Одержимо новий план x^2 ={7; 0; 10,3; 0; 1,7}. Як видно з таблиці, рядок оцінок містить тільки невід'ємні значення, тому досягнутий базис {1, 3, 5} є оптимальним. Отже, вектор x^*={7; 0; 10,3; 0; 1,7} є оптимальним планом (точкою максимуму) задачі, максимальне значення цільової функції дорівнює $L^* = L(x^*) = 360,8$.

Збіжність симплекс-методу. Виродженість у задачах ЛП. Найважливішою властивістю будь-якого обчислювального алгоритму є збіжність, тобто можливість одержання в ході його застосування шуканих результатів (із заданою точністю) за кінцеве число кроків (ітерацій).

Легко помітити, що проблеми зі збіжністю симплекс-методу потенційно можуть виникнути на етапі вибору значення r у випадку, коли однакові мінімальні значення відношення $\Theta_r = \min\limits_i \left\{ \frac{b_i}{a_{ij}} \right\}$ будуть досягнуті для кількох рядків таблиці одночасно. Тоді на наступній ітерації стовпець b^{q+1}-й міститиме нульові елементи. Нагадаємо, що припустимий базисний план канонічної задачі ЛП, що відповідає поточному базису, називають виродженим, якщо деякі його базисні компоненти дорівнюють нулю, тобто вектор b^q містить нульові елементи.

Задачу ЛП, що має вироджені плани, називають виродженою. При виході на вироджений план ми фактично одержуємо розкладання вектора $\overline{B}$ на меншу, ніж m, кількість векторів $\overline{A}_j$ і, отже, втрачаємо можливості коректно визначити, введення якого стовпця до базису приведе до зростання значення цільової функції. Подібні ситуації, в остаточному підсумку, можуть призвести до зациклення обчислювального процесу, тобто до нескінченного перебору тих самих базисів.

З погляду першої геометричної інтерпретації ЗЛП ситуація виродженості означає, що через певну кутову точку багатогранної множини припустимих планів задачі, що відповідає поточному базисному плану, проходить більш за m

гіперплощин обмежень задачі. Іншими словами, одне або кілька обмежень у певній точці є надлишковими. Останнє дає повід для міркувань про економічну сторону проблеми виродженості як проблеми наявності надлишкових обмежень і в певних випадках визначає шляхи її розв'язання.

Ідея методу розв'язання вироджених задач ЛП, що дістала назви **методу збурень**, полягає в тому, що при виході на вироджений план здійснюють незначний зсув вектора $\overline{B}$, і виродженість усувається.

Необхідно сказати, що розглянута проблема зациклення для більшості практично значущих задач є досить рідкою й мало імовірною. Більш того, вона дуже часто вирішується за рахунок помилок округлень при виконанні розрахунків на ЕОМ.

Знаходження припустимого базисного плану.У розглянутому вище прикладі вихідний базисний план, необхідний для початку обчислень за симплекс-методом, був підібраний за рахунок особливостей матриці умов. Дійсно, ця матриця вже містила необхідну кількість «майже базисних» стовпців. Очевидно, що для переважної більшості задач ЛП неможливо подібним чином відразу й у явному вигляді вказати вихідний припустимий базисний план. Існують різні прийоми розв'язання цієї задачі. Ми зупинимося на одному з них, що отримав назву **методу мінімізації нев'язань**. Його сильною стороною, безумовно, є універсальність. Хоча, у певних окремих випадках, він може опинитися занадто громіздким.

Ідея методу мінімізації нев'язань полягає в побудові відповідної допоміжної задачі, для якої можна в явному вигляді вказати вихідний базисний план, і розв'язанні її за допомогою процедури симплекс-методу.

Можна вважати, що вектор обмежень у вихідній задачі невід’ємний, тобто $b_i > 0, \quad i = \overline{1,m}$. Тоді для КЗЛП максимізації функції

$$L = c_1x_1 + c_2x_2 + \ldots + c_nx_n \to \max \tag{3.15}$$

на множині, зумовленій обмеженнями

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n &= b_1, \\ a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n &= b_2, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mn}x_n &= b_m, \\ x_j \ge 0, \quad j &= \overline{1,n} \end{aligned} \tag{3.16}$$

будують допоміжну задачу

$$L = 0x_1 + 0x_2 + \ldots + 0x_n - 1\, x_{n+1} - \ldots - 0 \cdot x_{n+m} \tag{3.17}$$

на новій множині $D' = \{x \in R^{n+m}\}$, зумовленій системою обмежень

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n + 1 * x_{n+1} + \ldots + 0 * x_{n+m} &= b_1, \\ a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n + 0 * x_{n+1} + 1 * x_{n+2} + \ldots + 0 * x_{n+m} &= b_2, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mn}x_n + 0 * x_{n+1} + 0 * x_{n+2} + \ldots + 1 * x_{n+m} &= b_m \\ x_j \ge 0, \quad j &= \overline{1, n+m}. \end{aligned} \tag{3.18}$$

Як видно, допоміжна задача (3.17)-(3.18) відрізняється від основної задачі (3.15)-(3.16) матрицею, що отримана в результаті додавання до вихідної матриці $\overline{A}$ одиничної матриці. Їй відповідають додаткові (фіктивні) змінні $x_{n+1}, ..., x_{n+m}$. Ці змінні (їх називають нев'язаннями) у цільовій функції мають коефіцієнти (-1), а змінні $x_1, ..., x_n$, які є спільними для обох задач, входять до цільової функції з нульовими коефіцієнтами. Істотною особливістю допоміжної задачі є наявність в неї очевидного вихідного базису, утвореного доданими стовпцями, і відповідного плану з базисними компонентами $x_{n+i}=b_i \geq 0$, $i=\overline{1,m}$. В силу структури цільової функції в процесі пошуку її максимуму за процедурою симплекс-методу фіктивні змінні (нев'язання) x_{n+i} повинні мінімізуватися, бажано до нуля, чого можна досягнути за рахунок послідовного переведення їх у небазисні компоненти плану. Якщо в оптимальному плані x^*, отриманому в результаті розв'язання допоміжної задачі, останні m компонент (тобто нев'язання) дорівнюють нулю, то його початкові n компонент задовольняють системі обмежень, які визначають область D, і x^* простим відкиданням нев'язань перетворюють на припустимий план основної задачі. При цьому всі рядки симплекс-таблиці, отриманої на останній ітерації допоміжної задачі, переносять до першої таблиці основної задачі.

У випадку, коли оптимальний план допоміжної задачі все-таки містить фіктивні компоненти, що не дорівнюють нулю, припустимі плани у вихідної задачі відсутні, тобто $D = \varnothing$.

Контрольні запитання

1. Сформулюйте задачу лінійного програмування.
2. Дайте визначення для наступних понять: план, припустимий план, оптимальний план, розв'язок задачі.
3. Поясніть, чим відрізняється загальна задача лінійного програмування від канонічної?
4. Чи завжди загальну задачу лінійного програмування можна привести до канонічного виду?
5. Яку точка опуклої множини називають кутовою?
6. У чому полягає перша геометрична інтерпретація задачі лінійного програмування?
7. Який план ЗЛП називають базисним?
8. Як пов'язані базисні плани й кутові точки області визначення задачі лінійного програмування?
9. Який план задачі лінійного програмування називають виродженим?
10. Сформулюйте критерій оптимальності припустимого базисного плану, застосовуваний у симплекс-методі.
11. Сформулюйте основні етапи стандартної ітерації симплекс-методу.
12. Для чого застосовують перетворення Жордана-Гауса?
13. Який елемент симплекс-таблиці називають ведучим?

14. При яких умовах роблять висновок про необмеженість цільової функції в розв'язуваній задачі?
15. Чи можна заздалегідь точно визначити кількість ітерацій, що необхідна для розв'язання задачі за симплекс-методом? Чи можна знайти верхню границю для даної величини?
16. Яку задачу називають виродженою? За якими ознаками можна впізнати, що поточний план є виродженим?
17. Поясніть, в чому полягає основна ідея методу збурень?
18. Для чого призначений метод мінімізації нев'язань? Поясніть сутність цього методу.

ТЕМА 4
ТЕОРІЯ ДВОЇСТОСТІ І ДВОЇСТІ ОЦІНКИ В АНАЛІЗІ РОЗВ’ЯЗКІВ ЛІНІЙНИХ ОПТИМІЗАЦІЙНИХ МОДЕЛЕЙ

4.1 Пряма і двоїста задачі як пара сполучених задач ЛП

Більшість задач лінійного програмування можна розглядати як економічні задачі про розподіл ресурсів $b_1, b_2, ..., b_m$ між різними споживачами, зокрема, між технологічними процесами $\overline{A}_1, \overline{A}_2, ..., \overline{A}_n$.

Будь-який припустимий розв’язок задачі лінійного програмування $x_1, x_2, ..., x_n$ дає конкретний розподіл ресурсів, що вказує частку кожного з них, яка буде витраченою при реалізації певного технологічного процесу.

Загальна задача лінійного програмування має вигляд:

знайти

$$L = c_1 x_1 + c_2 x_2 + ... + c_n x_n \to \max, \tag{4.1}$$

де x задовольняє обмеженням

$$\begin{aligned} &a_{11}x_1 + a_{12}x_2 + ... + a_{1n}x_n \le b_1, \\ &a_{21}x_1 + a_{22}x_2 + ... + a_{2n}x_n \le b_2, \\ &\dots\dots\dots\dots\dots\dots\dots\dots \\ &a_{m1}x_1 + a_{m2}x_2 + ... + a_{mn}x_n \le b_m, \\ &x_j \ge 0. \end{aligned} \tag{4.2}$$

Тут a_{ij} – витрати i-го ресурсу на j-й технологічний процес. Зміст обмежень – сумарні витрати i-го ресурсу не повинні перевищувати кількість цього ресурсу, наявного в нашому розпорядженні b_i.

Для будь-якого технологічного процесу актуальним є також питання про те, щоб витрати ресурсів були мінімальними. Позначимо u_i ціну одиниці i-го ресурсу. Тоді можна записати вимогу, щоб вартість ресурсів була мінімальною

$$L' = b_1 u_1 + b_2 u_2 + ... + b_m u_m \to \min,$$

а величина витрат на кожний вид технологічного процесу повинна бути не меншою за його вартість

$$a_{11}u_1 + a_{12}u_2 + ... + a_{1m}u_m \ge c_1,$$
$$a_{12}u_1 + a_{22}u_2 + ... + a_{m2}u_m \ge c_2,$$
$$\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots$$
$$a_{1n}u_1 + a_{2n}u_2 + ... + a_{mn}u_m \ge c_n,$$
$$u_i \ge 0.$$

Отже, одержали задачу:

знайти такий розв’язок u^*, що перетворює на мінімум цільову функцію

$$L' = \sum_{i=1}^{m} b_i u_i \to \min \tag{4.3}$$

і задовольняє обмеженням

$$\begin{aligned} &a_{11}u_1 + a_{12}u_2 + ... + a_{1m}u_m \ge c_1, \\ &a_{12}u_1 + a_{22}u_2 + ... + a_{m2}u_m \ge c_2, \\ &\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots \\ &a_{1n}u_1 + a_{2n}u_2 + ... + a_{mn}u_m \ge c_n, \\ &u_i \ge 0. \end{aligned} \tag{4.4}$$

Цю задачу називають **двоїстою** стосовно вихідної задачі, яку називають **прямою** задачею.

Порівняємо умови прямої та двоїстої задач у матричній формі:

Пряма	Двоїста
Знайти таке $\overline{X}^*$, що	Знайти таке $\overline{U}^*$, що
$\overline{C}^T\,\overline{X}^* \to \max$	$\overline{B}^T\,\overline{U}^* \to \min$
при обмеженнях	при обмеженнях
$\overline{A}\overline{X} \le \overline{B}, \quad \overline{X} \ge 0$.	$\overline{A}^T\,\overline{U} \ge \overline{C}, \quad \overline{U} \ge 0$.

Загальна схема побудови двоїстої задачі. Якщо задано загальну задачу ЛП, де множина припустимих планів D визначається системою рівнянь і нерівностей (4.2), то двоїстою щодо неї називається загальна задача ЛП, де D' визначається системою рівнянь і нерівностей (4.4).

При переході від прямої задачі ЛП до двоїстої:

- тип оптимуму змінюється на протилежний, тобто якщо пряма задача є задачею максимізації, то двоїста буде задачею мінімізації й навпаки;

- вектор коефіцієнтів цільової функції $\overline{C}$ й стовпець обмежень $\overline{B}$ міняються місцями. Тобто коефіцієнти цільової функції прямої задачі c_1, c_2, ..., c_n стають вільними членами обмежень двоїстої задачі, а вільні члени обмежень прямої задачі b_1, b_2, ..., b_m стають коефіцієнтами цільової функції двоїстої задачі;

- матрицю обмежень двоїстої задачі одержують шляхом транспонування матриці обмежень прямої задачі $\overline{A}$;

- множину індексів змінних, на які накладено умову невід’ємності в прямій задачі (наприклад, $x_j \ge 0$ або $u_i \ge 0$), визначають номера обмежень, що мають форму нерівностей у двоїстій задачі ($a_j u \ge c_j$ або $a_i x \le b_i$);

- множину номерів обмежень, що мають форму нерівностей у прямій задачі (наприклад, $a_i x \le b_i$ або $a_j u \ge c_j$), визначає множина індексів змінних, на які накладено умову невід'ємності, у двоїстій задачі ($u_i \ge 0$ або $x_j \ge 0$).

З наведених властивостей пари задач випливає важлива властивість - **симетричність відносини подвійності**, тобто задача, двоїста стосовно двоїстої, збігається з прямою (вихідною) задачею.

Тим самим можна говорити про пару взаємно двоїстих задач.

4.2 Основні теореми двоїстості, їх економічний зміст

Фундаментальні властивості двоїстих задач лінійного програмування формулюють у вигляді стверджень, що приводяться нижче. Їх зазвичай називають теоремами подвійності.

Теорема 4.1. (перша теорема подвійності). Якщо $\overline{X}_0$ і $\overline{U}_0$ - припустимі плани для пари двоїстих задач, тобто якщо

$$\overline{A}\ \overline{X}_0 \le \overline{B} \quad \text{і} \quad \overline{A}^{\mathrm{T}}\ \overline{U}_0 \ge \overline{C},$$

то

$$\overline{C}^{\mathrm{T}}\ \overline{X}_0 \le \overline{B}^{\mathrm{T}}\ \overline{U}_0,$$

тобто значення цільової функції прямої задачі ніколи не перевищують значень цільової функції двоїстої задачі.

Доказ.

Оскільки $\overline{U}_0$ - припустимий план, то

$$\overline{A}^{\mathrm{T}}\ \overline{U}_0 \ge \overline{C}; \tag{4.5}$$

оскільки $\overline{X}_0$ - припустимий план, то

$$\overline{A}\ \overline{X}_0 \le \overline{B}. \tag{4.6}$$

Помножимо (4.5) на $\overline{X}_0^T$

$$\overline{X}_0^T \overline{A}^T \overline{U}_0 \ge \overline{X}_0^T \overline{C}; \tag{4.7}$$

помножимо (4.6) на $\overline{U}_0^T$

$$\overline{X}_0 \overline{A}\overline{U}_0^T \le \overline{U}_0^T \overline{B} \tag{4.8}$$

і порівняємо (4.7) і (4.8). Оскільки

$$\overline{X}_0 \overline{A}\overline{U}_0^T = \left(\overline{X}_0^T \overline{A}^T \overline{U}_0\right)^T, \qquad \text{то } \overline{U}_0^T \overline{B} \ge \overline{X}_0^T \overline{C},$$

або

$$\overline{U}_0 \overline{B}^T \ge \overline{X}_0 \overline{C}^T.$$

Зауваження. Теорема 4.1, зрозуміло, вірна й для оптимальних планів взаємно двоїстих задач: $\overline{U}*\overline{B}^T \ge \overline{X}*\overline{C}^T$, де $\overline{X}$* і $\overline{U}$* - будь-які оптимальні плани задач. Насправді, як буде видно з подальшого, справедлива рівність $\overline{U}*\overline{B}^T = \overline{X}*\overline{C}^T$.

Теорема 4.2. (друга теорема подвійності). Якщо для певних припустимих планів $\overline{X}_0$ і $\overline{U}_0$ взаємно двоїстих задач виконується рівність

$$\overline{C}^{\mathrm{T}}\ \overline{X}_0 = \overline{B}^{\mathrm{T}}\ \overline{U}_0,$$

то $\overline{X}_0$ і $\overline{U}_0$ є оптимальними планами цих задач.

Доказ.

Відповідно до теореми 4 для всіх припустимих розв’язків $\overline{X}$ і $\overline{U}$ справедлива нерівність

$$\overline{C}^T \overline{X} \le \overline{B}^T \overline{U},$$

але оскільки з умови теореми

$$\overline{C}^T \overline{X}_0 = \overline{B}^T \overline{U}_0,$$

то $\overline{C}^T \overline{X}_0$ – найбільше з можливих значень цільової функції, тобто

$$\overline{C}^T \overline{X} \le \overline{C}^T \overline{X}_0$$

цільова функція від $\overline{X}_0$ – найбільша, Отже, $\overline{X}_0$ є оптимальним значенням.

Аналогічно для цільової функції двоїстої задачі

$$\overline{B}^T \overline{U}_0 \le \overline{B}^T \overline{U},$$

тобто $\overline{B}^T \overline{U}_0$ – найменше з можливих значень цільової функції двоїстої задачі, отже $\overline{U}_0$ – оптимальне значення.

Отже, якщо для певних припустимих планів $\overline{X}_0$ і $\overline{U}_0$ прямої і двоїстої задач їх цільові функції є рівними, то $\overline{X}_0$ й $\overline{U}_0$ – оптимальні плани пари сполучених задач.

Теорема 4.3. Якщо цільова функція в прямої задачі не обмежена зверху, то двоїста до неї задача не має припустимих планів.

Доказ.

Якщо припустити, що у двоїстій задачі існує хоча б один припустимий план $\overline{U}_0$, то відповідно до теореми 4.2, для будь-якого припустимого плану $\overline{X}_0$ прямої задачі справедлива нерівність $\overline{C}^T \overline{X} \le \overline{B}^T \overline{U} < +\infty$. Останнє означає, що цільова функція прямої задачі обмежена зверху. Оскільки це суперечить умові теореми, припущення про існування припустимих планів двоїстої задачі є невірним.

Наступне ствердження, відоме як теорема рівноваги, використовують при перевірці оптимальності планів ЗЛП.

Теорема 4.4. Нехай $\overline{X}^*$ і $\overline{U}^*$ - оптимальні плани канонічної та двоїстої щодо неї задач. Якщо j-та компонента плану $\overline{X}^*$ строго додатна ($x^*_j>0$), то відповідне j-е обмеження двоїстої задачі виконується як рівність: $a_{1j}u_1+\ldots+a_{mj}u_m=c_j$; якщо j-та компонента плану $\overline{X}^*$ має нульове значення ($x^*_j = 0$), то j-е обмеження двоїстої задачі виконується як нерівність: $a_{1j}u_1+\ldots+a_{mj}u_m>c_j$.

Доказ.

Вектори $\overline{X}^*$ і $\overline{U}^*$ як припустимі плани, задовольняють обмеженням

$$\overline{A}\overline{X}^* = \overline{B}, \qquad \overline{X}^* > 0 \qquad \text{- прямої задачі;}$$
$$\overline{A}^T\overline{U}^* - \overline{C}^T \geq 0, \qquad \overline{U}^* > 0 \text{ - двоїстої задачі.}$$

Запишемо скалярний добуток

$$\left(\overline{A}^T\overline{U}^* - \overline{C}^T\right)\overline{X}^* = \overline{A}^T\overline{X}^*\overline{U}^* - \overline{C}^T\overline{X}^* = \overline{B}^T\overline{U}^* - \overline{C}^T\overline{X}^*.$$

Одержали різницю цільових функцій прямої і двоїстої задач. На підставі другої теореми подвійності оптимальні значення цільових функцій взаємно двоїстих задач збігаються. Отже скалярний добуток

$$\left(\overline{A}^T\overline{U}^* - \overline{C}^T\right)\overline{X}^* = 0.$$

Але скалярний добуток двох невід'ємних векторів може дорівнювати нулю тільки в тому випадку, якщо всі попарні добутки їх відповідних координат дорівнюють нулю. Тоді, якщо $x_j > 0$, то $\sum_i a_{ij}u_i - c_j = 0$ або $\sum_i a_{ij}u_i = c_j$. А якщо $x_j = 0$, то можливо, що $\sum_i a_{ij}u_i - c_j > 0$ або $\sum_i a_{ij}u_i > c_j$.

Практичне значення теорем подвійності полягає в тому, що вони дозволяють замінити процес розв'язання прямої задачі на розв'язання двоїстої, який в певних випадках може виявитися простішим. Наприклад, задачу, область припустимих значень якої описується двома рівняннями, що пов'язують шість змінних ($m = 2$, $n = 6$), не можна вирішити графічним методом. Проте цей метод можна застосувати для розв'язання двоїстої до неї задачі, яка має тільки дві змінні.

Оптимальний розв'язок двоїстої задачі можна одержати з таблиці, отриманої на фінальній ітерації процедури симплекс-методу. Елементи індексного рядка цієї таблиці L_j обчислюють відповідно до виразу (3.9) $L_j = \sum_{i=1}^{m} c_j a_{ij}$, де c_j – елементи вектора-рядка, що містить коефіцієнти цільової функції прямої задачі при змінних, які є базисними в оптимальному плані, а a_{ij} - елементи матриці $\overline{D}^{-1}$, що зворотна до матриці $\overline{D}$. Матриця $\overline{D}$ складається з базисних векторів оптимального плану, компоненти яких узяті з початкового опорного плану задачі. Оптимальний план двоїстої задачі визначається співвідношенням

$$\overline{U}^* = \overline{c}_{j\text{баз}}\overline{D}^{-1}. \tag{4.9}$$

Зворотна матриця $\overline{D}^{-1}$ завжди міститься в останній симплекс-таблиці у тих стовпчиках, де в першій таблиці задачі перебувала одинична матриця. (Нагадаємо, що добуток матриці на її зворотну дає одиничну матрицю, в якої діагональні елементи дорівнюють 1, а всі інші дорівнюють 0).

Отже, зв'язок між оптимальними розв'язками прямої й двоїстої задач і елементами індексних рядків L_j симплекс-таблиць, які відповідають цим розв'язкам, виражається наступними співвідношеннями:

$$a_{0,n+i}^{ПР} = u_i^*, \quad i = \overline{1,m},$$

$$-a_{0,m+j}^{ДВ} = x_j*, \quad j = \overline{1,n},$$

де n – кількість змінних прямої задачі;

m – кількість обмежень прямої задачі;

$a_{0,n+i}^{ПР}$ - $(n+i)$-й елемент індексного рядка симплекс-таблиці прямої задачі, що містить оптимальний план; $a_{0,m+j}^{ДВ}$ - $(m+j)$-й елемент індексного рядка симплекс-таблиці двоїстої задачі, що містить оптимальний план.

4.3 Двоїсті оцінки і дефіцитність ресурсів

У різних джерелах компоненти оптимального плану двоїстої задачі називають **двоїстими оцінками** або **тіньовими цінами**.

На підставі теорем подвійності для пари задач ЛП у загальній формі формулюють певні важливі з погляду економічної інтерпретації слідства. Зокрема, з теореми 4.4 випливає, що якщо при реалізації оптимального плану прямої задачі і-е обмеження виконується як строга нерівність, то оптимальне значення відповідної двоїстої змінної дорівнює нулю:

$$a_{i1}x_1^* + \cdots + a_{in}x_n^* < b_i, \text{ то } u_i^* = 0.$$

Це означає, що якщо певний ресурс b_i є в надлишковій кількості (не витрачається повністю при реалізації оптимального плану), то i-е обмеження стає несуттєвим, і тіньова оцінка такого ресурсу дорівнює нулю. Отже, тіньові оцінки характеризують **дефіцитність ресурсів**.

Якщо при реалізації оптимального плану двоїстої задачі j-е обмеження виконується як строга нерівність, то оптимальне значення відповідної змінної в оптимальному плані прямої задачі повинне дорівнювати нулю

$$a_{1j}u_1^* + \cdots + a_{mj}u_m^* > c_j, \text{ то } x_j^* = 0.$$

Цей факт виражає **принцип рентабельності виробництва**. З огляду на економічний зміст двоїстих оцінок $u_1^*, \cdots, u_m^*$, вираз $a_{1j}u_1^* + \cdots + a_{mj}u_m^*$ можна інтерпретувати як питомі витрати на j-й технологічний процес. Отже, якщо ці витрати перевищують прибуток від реалізації одиниці j-го продукту, виробництво j-го продукту є нерентабельним і не повинне бути присутнім в оптимальному виробничому плані ($x_j* = 0$).

Незважаючи на можливі аналогії, які можуть виникнути у зв'язку з такими фундаментальними поняттями економічної теорії, як граничні витрати й граничний дохід, двоїсті оцінки не можна однозначно ототожнювати із цінами (хоча такі спроби іноді вживали на початковій стадії становлення дослідження операцій як науки).

Контрольні запитання

1. Поясніть сутність подвійності в лінійному програмуванні.
2. Складіть просту економіко-математичну модель і запишіть до неї двоїсту. Дайте економічну інтерпретацію двоїстих оцінок.

3. Скільки змінних і обмежень має двоїста задача стосовно прямої задачі?
4. Поясніть економічний зміст першої теореми подвійності.
5. Поясніть економічний зміст другої теореми подвійності.
6. У чому полягає економічний зміст третьої теореми подвійності?
7. Сформулюйте правила побудови двоїстих задач.
8. Як на підставі оптимального розв’язку прямої задачі одержати оптимальний розв’язок двоїстої задачі?

ТЕМА 5
АНАЛІЗ ЛІНІЙНИХ МОДЕЛЕЙ ЕКОНОМІЧНИХ ЗАДАЧ

5.1 Аналіз розв’язків лінійних економіко-математичних моделей

Традиційна економічна інтерпретація двоїстої задачі ЛП базується на моделі найпростішої задачі виробничого планування. В неї кожний j-й елемент вектора $\overline{X}$ розглядається як план випуску продукції певного виду в натуральних одиницях, c_j - ціна одиниці продукції j-го виду, $\overline{A}_j$ - вектор, що визначає технологію витрати наявних m ресурсів на виробництво одиниці продукції j-го виду, $\overline{B}$ - вектор обмежень на обсяги цих ресурсів.

Припустимо, що для певних значень $\overline{A}_j$, $\overline{B}$ і c_j знайдений оптимальний план x^* прямої задачі, що максимізує сумарний дохід $\max\limits_{x \in D}\{cx\} = cx^*$, і визначені оптимальні оцінки сировини, тобто оптимальний план двоїстої задачі u^*. З виразу цільової функції двоїстої задачі $L'^* = b_1 u_1^* + b_2 u_2^* + \ldots + b_m u_m^*$ видно, що величина двоїстої оцінки u_i^* показує, наскільки зросте максимальне значення цільової функції прямої задачі при збільшенні кількості сировини відповідного виду на одну одиницю. Отже, змінні двоїстої задачі $u_1^*, \cdots, u_m^*$ за своїм змістом є оцінками потенційної можливості одержання додаткового прибутку за рахунок збільшення обсягу відповідного ресурсу в умовах оптимального функціонування керованого економічного об'єкта.

Виникає питання про те, як змінюватиметься оптимальний план x^* при зміні компонентів вектора обмежень $\overline{B}$ і, зокрема, при яких варіаціях $\overline{B}$ оптимальний план x^* залишиться оптимальним. Ця задача одержала назву проблеми стійкості оптимального плану. Очевидно, що дослідження стійкості x^* має й практичне значення, оскільки в реальному виробництві обсяги доступних ресурсів b_i можуть істотно коливатися після ухвалення планового розв’язку x^*.

Коли вектор обмежень $\overline{B}$ отримує приріст Δb, виникають відповідні варіації для оптимального плану прямої задачі $x^*(b+\Delta b)$ і значення цільової функції $L[x^*(b+\Delta b)]$. Припустимо, приріст Δb такий, що він не призводить до зміни оптимального базису задачі, тобто $x^*(b+\Delta b) \geq 0$. Введемо функцію $F(b)$, що повертає оптимальне значення цільової функції задачі для різних значень вектора обмежень $\overline{B}$

$$F(b) = \max_{x \in D(b)} \{cx\}. \tag{5.1}$$

Розглянемо відношення її приросту $F(b+\Delta b)$-F(b) до приросту аргументу Δb. Якщо для певного і спрямувати $\Delta b \to 0$, одержимо

$$\lim_{\Delta b_i \to 0} \frac{F(b+\Delta b) - F(b)}{\Delta b_i} = \frac{\partial F(b)}{\partial b_i}. \tag{5.2}$$

З огляду на те, що відповідно до теореми 4.2 цільові функції пари сполучених задач при оптимальних планах рівні між собою, запишемо

$$F(b) = \sum_{j=1}^{n} c_j x_j^* = \sum_{i=1}^{m} b_i u_i^*. \tag{5.3}$$

Підставимо (5.3) до (5.2) і одержимо вираз

$$\frac{\partial F(b)}{\partial b_i} = \frac{\partial \left(\sum_{i=1}^{m} b_i u_i^* \right)}{\partial b_i} = u_i^*. \tag{5.4}$$

Теорема 5.1. В оптимальному плані двоїстої задачі значення змінної u_i* чисельно дорівнює частинній похідній цільової функції *L'** за відповідним аргументом b_i.

Звідси випливає економічна інтерпретація оптимальних змінних двоїстої задачі:

Кожний елемент $u_1^*, \cdots, u_m^*$ може розглядатися як гранична (миттєва) оцінка внеску і-го ресурсу в сумарний дохід *L** при оптимальному розв'язку $x_1^*, \cdots, x_n^*$.

Інакше кажучи, u_i^* дорівнює приросту доходу, що виникає при збільшенні ресурсу і на одиницю за умови оптимального використання ресурсів.

Отже, якщо знайдено оптимальний план прямої задачі, можна провести аналіз стійкості двоїстих оцінок щодо змін компонентів вектора $\overline{B}$. Це дозволяє оцінити стійкість оптимального плану двоїстої задачі щодо зміни обмежень прямої задачі й ступінь впливу зміни $\overline{B}$ на максимальне значення цільової функції, а також визначити найдоцільніший варіант можливих змін $\overline{B}$.

План двоїстої задачі не змінюється для всіх значень $b_i + \Delta b_i$, при яких стовпець вектора $\overline{B}*$ останньої симплекс-таблиці не містить від'ємних чисел, тобто коли серед компонентів вектора

$$\overline{B}* = \begin{vmatrix} b_1 + \Delta b_1 \\ b_2 + \Delta b_2 \\ \cdots\cdots\cdots \\ b_m + \Delta b_m \end{vmatrix} \tag{5.5}$$

немає від'ємних. $\overline{B}*$ – матриця, складена з компонентів векторів базису, що визначає оптимальний план задачі, оскільки базисні компоненти оптимального плану перебувають у стовпці вектора $\overline{B}$ останньої симплекс-таблиці.

Елементи (n+і)-го стовпця a_{ij} останньої симплекс-таблиці, що містить оптимальну оцінку і-го ресурсу u_i*, показують, на скільки одиниць змінюються компоненти оптимального плану $x*$ при збільшенні обсягу цього ресурсу на одиницю, тобто $x_j*(b_i+\Delta b_i)=x_j*(b_i)+a_{i,n+i}\Delta b_i$. Припустимі інтервали зміни для і-го ресурсу можна визначити з умови:

$$\overline{B}*=\begin{vmatrix} x_1+a_{1,n+i}\Delta b_i\ge 0 \\ x_2+a_{2,n+i}\Delta b_i\ge 0 \\ \dots\dots\dots\dots\dots\dots \\ x_n+a_{n,n+i}\Delta b_i\ge 0 \end{vmatrix}. \qquad (5.6)$$

Розглянемо приклад. Нехай остання симплекс-таблиця, що містить оптимальний план, має вигляд, наведений у таблиці 5.1.

Таблиця 5.1 - Симплекс-таблиця, що містить оптимальний план

Базис	$C_{j\text{баз}}$	C_j	6	5	0	0	0
		B	A_1	A_2	A_3	A_4	A_5
A_1	6	83	1	0	4/7	-3/7	0
A_2	5	36	0	1	-3/7	4/7	0
A_5	0	21	0	0	11/28	-11/7	1
індексний рядок	L_j	678	6	5	9/7	2/7	0
	Δ_j		0	0	9/7	2/7	0

Звідки оптимальний план прямої задачі $x*$=(83; 36; 0; 0; 21), оптимальний план двоїстої задачі $u*=(\frac{9}{7}, \frac{2}{7}, 0)$. Дефіцитнішою є сировина S_1, оскільки її тіньова оцінка вища й відповідно сильніше впливає на величину цільової функції.

Очевидно, що збільшення доходу можна отримати тільки шляхом зміни оптимального плану прямої задачі. З таблиці видно, що при збільшенні на 1 одиницю кількості сировини S_1, дохід збільшиться на $\frac{9}{7}$ грн. Це відбудеться якщо виробництво виробів A збільшити на $\frac{4}{7}$ одиниці, а виробництво виробів B знизити на $\frac{3}{7}$ одиниці, при цьому витрата сировини S_3 зросте на $\frac{11}{28}$ одиниці. Новий оптимальний план прямої задачі матиме вигляд $x*'=(83\frac{4}{7}, 35\frac{4}{7}, 0, 0, 20\frac{17}{28})$, а прибуток становитиме $L*=6*83\frac{4}{7}+5*35\frac{4}{7}+0*0+0*20\frac{17}{28}=$ =679,286 грн.

Визначимо інтервали стійкості двоїстих оцінок. Для ресурсу 1 відповідно до елементів стовпчика A_3 маємо

$$x*=(83+0{,}57\Delta b_1;\, 36-0{,}43\Delta b_1;\, 0;\, 0;\, 21+0{,}39\Delta b_1).$$

Запишемо вектор $\overline{B}*$ з умовами його невід'ємності й визначимо межі припустимих значень Δb_1

$$\overline{B}* = \left| \begin{array}{l} 83+0{,}57\Delta b_1 \ge 0 \\ 36-0{,}43\Delta b_1 \ge 0 \\ 21+0{,}39\Delta b_1 \ge 0 \end{array} \right| = \left| \begin{array}{l} \Delta b_1 \ge -66{,}7 \\ \Delta b_1 \le 83{,}7 \\ \Delta b_1 \ge -53{,}8. \end{array} \right|$$

Оптимальний план двоїстої задачі залишиться незмінним, якщо Δb_1 належить інтервалу $-53{,}8 \le \Delta b_1 \le 83{,}7$, а перший ресурс $440-53{,}7 \le b_1 \le 440+83{,}7$ або $386{,}3 \le b_1 \le 523{,}7$.

Аналогічно для ресурсу 2 відповідно до елементів стовпчика A_4 запишемо

$$x* = (83-0{,}43\Delta b_2;\, 36+0{,}57\Delta b_2;\, 0;\, 0;\, 21-1{,}57\Delta b_2).$$

$$\overline{B}* = \left| \begin{array}{l} 83-0{,}43\Delta b_2 \ge 0 \\ 36+0{,}57\Delta b_2 \ge 0 \\ 21-1{,}57\Delta b_2 \ge 0 \end{array} \right| = \left| \begin{array}{l} \Delta b_2 \le 193 \\ \Delta b_2 \ge -63{,}2 \\ \Delta b_2 \le 13{,}4. \end{array} \right|$$

Оптимальний план двоїстої задачі залишиться незмінним, якщо Δb_2 належить інтервалу $-63{,}2 \le \Delta b_2 \le 13{,}4$, а другий ресурс $393-63{,}2 \le b_2 \le 393+13{,}4$ або $329{,}8 \le b_2 \le 406{,}4$.

Отже, якщо збільшення кількості ресурсів S_1 належить проміжку $-53{,}8<\Delta b_1<83{,}7$, а кількість інших ресурсів незмінна, або збільшення кількості ресурсів S_2 належить проміжку $-63{,}2<\Delta b_2<83{,}7$, а кількість інших ресурсів незмінна, то двоїста задача має той самий оптимальний план $u* = (1{,}286;\, 0{,}286;\, 0)$.

Стосовно прямої задачі, можна показати, що при зміні кількості першого ресурсу S_1 у межах $386{,}3 \le b_1 \le 523{,}7$ можливий дохід підприємства лежить у межах $609{,}7 \le L* \le 784{,}3$ а оптимальний план прямої задачі

$$(52{,}3;\, 59{,}1;\, 0;\, 0;\, 0{,}02) \le x* \le (131;\, 0;\, 0;\, 0;\, 53{,}6).$$

При зміні кількості другого ресурсу S_2 у межах $329{,}8 \le b_2 \le 406{,}4$ можливий дохід підприємства лежить у межах $661 \le L* \le 681{,}6$ а оптимальний план прямої задачі

$$(110;\, 0;\, 0;\, 0;\, 53{,}6) \le x* \le (77;\, 43{,}6;\, 0;\, 0;\, 0).$$

Розраховані інтервали належать випадкам, коли змінюється тільки один ресурс. У випадку одночасної зміни обсягу всіх або кількох ресурсів для визначення нового оптимального плану використовують одне з основних співвідношень обчислювальної процедури симплекс-методу:

$$x* = \overline{D}^{-1} * \overline{B}, \tag{5.7}$$

де $\overline{D}$ - матриця, що складається з базисних векторів оптимального плану, компоненти яких узяті з початкового опорного плану;

$\overline{B}$ - вектор обмежень.

У розглянутому числовому прикладі матриці $\overline{D}$ й $\overline{D}^{-1}$ відповідно мають вигляд

$$\overline{D} = \left| \begin{array}{ccc} 4 & 3 & 0 \\ 3 & 4 & 0 \\ 3 & 5 & 1 \end{array} \right|; \qquad \overline{D}^{-1} = \left| \begin{array}{ccc} 4/7 & -3/7 & 0 \\ -3/7 & 4/7 & 0 \\ 11/28 & -11/7 & 1 \end{array} \right|$$

Нехай новий вектор обмежень

$$\overline{B} = \begin{vmatrix} 440+84 \\ 393+13,4 \\ 450+0 \end{vmatrix} = \begin{vmatrix} 524 \\ 406,4 \\ 450 \end{vmatrix};$$

тоді новий оптимальний план визначиться в такий спосіб

$$x^* = \begin{vmatrix} 4/7 & -3/7 & 0 \\ -3/7 & 4/7 & 0 \\ 11/28 & -11/7 & 1 \end{vmatrix} * \begin{vmatrix} 524 \\ 406,4 \\ 450 \end{vmatrix} = \begin{vmatrix} 125,26 \\ 7,66 \\ 17,2. \end{vmatrix}$$

Тобто x^*=(125,26; 7,66; 0; 0; 17,2), при якому прибуток дорівнюватиме 790 грош. од.

5.2 Аналіз параметричної стійкості розв'язків ЗЛП

З погляду економічної інтерпретації задачу з дослідження параметричної стійкості можна розглядати як вивчення тих меж коливання цін на продукцію керованого підприємства (фірми), при яких прийнятий план випуску продукції залишається оптимальним. Отже, питання стійкості оптимального плану ЗЛП може бути поставленим для випадку варіації коефіцієнтів цільової функції

$$c_j, \quad j = \overline{1,n}.$$

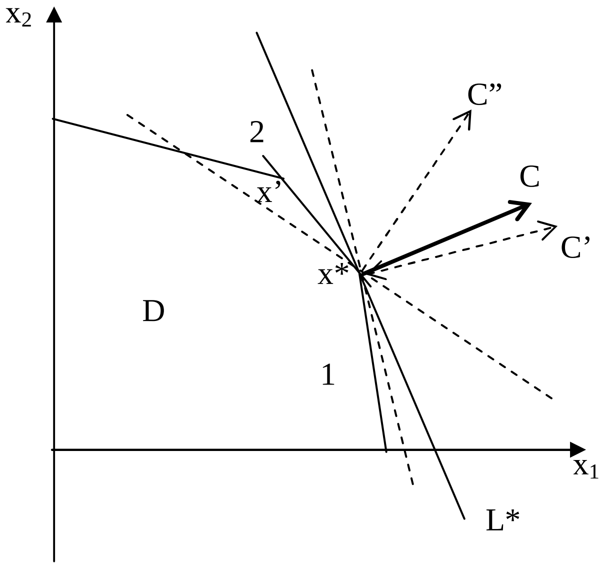


Рисунок 5.1 - Графічна інтерпретація параметричної стійкості оптимального плану

Зміст проблеми стійкості оптимального плану ЗЛП щодо варіацій цільової функції може бути проілюстрованим за допомогою першої геометричної інтерпретації. На рисунку 5.1 зображено множину припустимих планів *D* певної задачі ЛП. Як видно з рисунку, цільова функція *L* досягає екстремального значення в точці x^*, а зміні її коефіцієнтів від *c* до *c*' або *c*'' на рисунку відповідає поворот лінії рівня відносно x^*. Активним обмеженням (тобто таким, що звертаються на рівність) у точці x^* відповідають лінії 1 і 2. Доти, поки при повороті, що викликаний зміною вектора *c*, лінія рівня цільової функції не виходить за межі утворені лініями обмежень множини, x^* залишається оптимальним планом. Як показано на рисунку 5.1, цей план не змінюється при переході від *c* до *c*', і, навпаки, при переході від *c* до *c*'' лінія рівня цільової функції $L(x)=c''x$ перетинає лінію 2, що викличе зміну оптимального базисного плану, яким тепер стане точка *x*'. Використовуючи умови оптимальності плану ЗЛП

$$\Delta_j = L_j - c_j \geq 0, \tag{5.8}$$

можна одержати кількісні оцінки для меж коливань коефіцієнтів цільової функції, при яких не відбувається зміна оптимального плану. Припустимо, що варіації піддався певний елемент $c_r' = c_r + \Delta c_r$. Можливі два випадки:

1. Стовпець r не входить до оптимального базису. Тоді для незмінності оптимального плану необхідно й достатньо виконання умови

$$\Delta_r' = L_r - c_r' \geq 0.$$

Звідси можна одержати значення для припустимої варіації

$$\Delta c_r \leq L_r - c_r. \tag{5.9}$$

2. Стовпець r входить до оптимального базису. У цьому випадку для збереження оптимальності поточного плану буде потрібним виконання для всіх небазисних стовпців умов (5.8) або

$$\Delta_j = \sum_{i=1}^{m} c_{j\,баз} a_{ij} - c_j \geq 0, \tag{5.10}$$

оскільки у цьому випадку зміни відбуваються також у стовпчику $C_{баз}$ симплекс-таблиці, а це, у свою чергу, стосується всіх ненульових оцінок Δ_j.

Отже, у цьому випадку припустима варіація повинна задовольняти умовам

$$\Delta c_r \leq \sum_{i=1}^{m} c_{j\,баз} a_{ij} - c_j. \tag{5.11}$$

Повернемося до числового прикладу й визначимо межі зміни параметрів цільової функції, при яких знайдений план $x^*=(83; 36; 0; 0; 21)$ залишається оптимальним. У цій задачі інтерес представляють варіації коефіцієнтів c_1 і c_2, які стоять при базисних змінних в оптимальному плані.

Запишемо умови (5.11) для коефіцієнта c_1

$$\begin{cases} \Delta_3 = L_3 - c_3 = (6+\Delta c_1)*4/7 + 5*(-3/7) + 0*11/28 - 0 = \dfrac{24}{7} + \dfrac{4}{7}\Delta c_1 - \dfrac{15}{7} \\ \Delta_4 = L_4 - c_4 = (6+\Delta c_1)*(-3/7) + 5*4/7 + 0*11/7 - 0 = -\dfrac{18}{7} - \dfrac{3}{7}\Delta c_1 + \dfrac{20}{7} \end{cases} \Rightarrow$$

$$\begin{cases} \Delta_3 = \dfrac{24}{7} + \dfrac{4}{7}\Delta c_1 - \dfrac{15}{7} \geq 0 \\ \Delta_4 = -\dfrac{18}{7} - \dfrac{3}{7}\Delta c_1 + \dfrac{20}{7} \geq 0 \end{cases} \Rightarrow \begin{cases} \Delta c_1 \geq -\dfrac{9}{4} \\ \Delta c_1 \leq \dfrac{2}{3} \end{cases},$$

$$-\frac{9}{4} \leq \Delta c_1 \leq \frac{2}{3},$$

$$3{,}75 \leq c_1 \leq 6{,}67.$$

Аналогічно визначимо варіацію коефіцієнта c_2.

$$\begin{cases} \Delta_3 = L_3 - c_3 = 6*4/7 + (5+\Delta c_2)*(-3/7) + 0*11/28 - 0 = \dfrac{9}{7} - \dfrac{3}{7}\Delta c_2 \\ \Delta_4 = L_4 - c_4 = 6*(-3/7) + (5+\Delta c_2)*4/7 + 0*(-11/7) - 0 = \dfrac{2}{7} + \dfrac{4}{7}\Delta c_2 \end{cases} \Rightarrow$$

$$\begin{cases} \Delta_3 = \frac{9}{7} - \frac{3}{7}\Delta c_2 \ge 0 \\ \Delta_4 = -\frac{2}{7} + \frac{4}{7}\Delta c_2 \ge 0 \end{cases} \Rightarrow \begin{cases} \Delta c_2 \le 3 \\ \Delta c_2 \ge \frac{1}{2} \end{cases},$$

$$\frac{1}{2} \le \Delta c_2 \le 3,$$

$$5{,}5 \le c_2 \le 8.$$

Наведений приклад з дослідження чутливості оптимального плану щодо зміни параметрів задачі є простим. Існують і складніші задачі, в яких, наприклад, досліджують спільні варіації параметрів різних типів. Вони складають предмет спеціального розділу дослідження операцій, що одержав назву параметричного програмування.

5.3 Оцінка рентабельності виробленої продукції

Оцінка рентабельності продукції, що випускається підприємством, ґрунтується на теоремі рівноваги, (теоремі 4.4) і виконується за допомогою двоїстих оцінок і обмежень двоїстої задачі, що характеризують кожний вид продукції. З цієї теореми випливає, що кількість продукції, витрати на виробництво якої перевищують дохід, в оптимальному плані дорівнює нулю.

Якщо при підстановці u^* до системи обмежень двоїстої задачі вартість ресурсів, що витрачаються на одиницю продукції (ліва частина), перевищує ціну цієї продукції (права частина), то виробництво такої продукції для підприємства недоцільно. Тобто у цьому випадку цей вид продукції є нерентабельним. Якщо співвідношення виконується як рівність, продукція є рентабельною.

Аналогічні результати можна одержати, якщо проаналізувати симплекс-різниці Δ_j у стовпчиках, які відповідають досліджуваним видам продукції. Їх значення показують, наскільки вартість ресурсів перевищує ціну одиниці відповідної продукції. Так, якщо симплекс-різниця дорівнює нулю $\Delta_j = 0$, то продукція є рентабельною. І, навпаки, якщо $\Delta_j \ne 0$, то відповідна продукція нерентабельна. Помітимо також, що індексний рядок останньої симплекс-таблиці містить значення додаткових змінних двоїстої задачі u^*. Отже, якщо вони перевищують ціну продукції відповідного виду, то цей вид продукції є рентабельним.

Звернемося до числового прикладу. Підставимо отримані тіньові оцінки в обмеження двоїстої задачі:

$$\begin{cases} 4 * \frac{9}{7} + 3 * \frac{2}{7} + 3 * 0 = 6 & \textit{продукція A є рентабельною,} \\ 3 * \frac{9}{7} + 4 * \frac{2}{7} + 5 * 0 = 5 & \textit{продукція B є рентабельною.} \end{cases}$$

Оскільки обидва обмеження виконуються як строгі рівності, обидва види продукції *A* і *B* є рентабельними. Це підтверджується і тим, що в оптимальному плані x^*=(83; 36; 0; 0; 21) обидва елементи x_1 і x_2, що відповідають обсягам виробництва, не нульові.

Проаналізуємо додаткові змінні двоїстої задачі, які розміщаються в індексному рядку симплекс-таблиці в стовпчиках A_1 і A_2. Їх оптимальні значення u_4=6; u_5 = 5. Відповідно симплекс-різниці $\Delta_1 = \Delta_2 = 0$, що так само свідчить про рентабельність продукції А і В.

5.4 Аналіз обмежень дефіцитних і недефіцитних ресурсів

Статус ресурсів прямої задачі можна визначити трьома способами. Перший - підстановкою x^* до системи обмежень прямої задачі. Якщо обмеження виконується як рівність, то відповідний ресурс є дефіцитним, у протилежному випадку - недефіцитним.

Другий спосіб - за допомогою додаткових змінних прямої задачі. Якщо додаткова змінна в оптимальному плані дорівнює нулю, то відповідний ресурс дефіцитний, а якщо відмінна від нуля - ресурс недефіцитний.

Третій спосіб - за допомогою двоїстих оцінок. Якщо $u_i \neq 0$, то зміна (збільшення або зменшення) обсягів і-го ресурсу призводить до відповідної зміни доходу підприємства, і тому такий ресурс є дефіцитним. Якщо u_i=0, то і-й ресурс недефіцитний.

Звернемося до числового прикладу. Підставимо компоненти оптимального плану x^*=(83; 36; 0; 0; 21) до системи обмежень прямої задачі:

$$\begin{cases} 4*83+3*36=440 & \textit{ресурс витрачено повністю,} \\ 3*83+4*36=393 & \textit{ресурс витрачено повністю,} \\ 3*83+5*36=429 & \textit{ресурс витрачено не повністю.} \end{cases}$$

Змінні u_1 = 1,286 і u_2 = 0,286 є умовними двоїстими оцінками одиниці сировини S_1 і S_2 відповідно. Ці оцінки відмінні від нуля, а сировина S_1 і S_2 повністю витрачається при оптимальному плані виробництва виробів А і В. Двоїста оцінка одиниці сировини S_3 u_3=0. Цей вид сировини не витрачається повністю при оптимальному плані виробництва.

Отже, додатну двоїсту оцінку мають лише ті види сировини, які повністю витрачаються при оптимальному плані виробництва. Тому двоїсті оцінки визначають дефіцитність використовуваної підприємством сировини.

$$\begin{cases} u_1 = 1{,}286 & \textit{ресурс } 1 \textit{ дефіцитний,} \\ u_2 = 0{,}286 & \textit{ресурс } 2 \textit{ дефіцитний,} \\ u_3 = 0 & \textit{ресурс } 3 \textit{ недефіцитний.} \end{cases}$$

Контрольні запитання

1. У чому полягає економічна інтерпретація прямої й двоїстої задач лінійного програмування?
2. Як визначити, чи є ресурс дефіцитним?
3. Як визначити, що продукція є рентабельною або нерентабельною?
4. У чому полягає економічний зміст змінних двоїстої задачі?
5. Який зміст вкладають у поняття «параметрична стійкість»?

6. Сформулюйте умови для припустимих змін цільової функції задачі, при яких її оптимальний план залишається незмінним.

7. Як визначити статус ресурсів прямої задачі?

8. Як визначити інтервали стійкості двоїстих оцінок щодо зміни запасів дефіцитних ресурсів?

9. Як визначити оптимальний план виробництва продукції й зміну доходу підприємства при збільшенні або зменшенні обсягу ресурсів?

10. Як розрахувати інтервали можливої зміни ціни одиниці кожного виду продукції?

ТЕМА 6
ТРАНСПОРТНА ЗАДАЧА

6.1 Транспортна задача та її властивості

Економіко-математична модель задачі. Нехай є m пунктів відправлення (постачальників) A_1, A_2, ..., A_m, у яких знаходиться однорідна продукція в кількостях a_1, a_2, ..., a_m відповідно. Нехай є n пунктів призначення (споживачів) B_1, B_2, ..., B_n, що подали заявки відповідно на b_1, b_2, ..., b_n одиниць вантажу. Сума всіх заявок дорівнює сумі всіх запасів.

Відомі вартості c_{ij} перевезення одиниці продукції з і-го пункту відправлення в j-й пункт призначення. Вважається, що вартість перевезення кількох одиниць вантажу пропорційна їх числу.

Потрібно скласти такий план перевезень (звідки, куди й скільки одиниць везти), щоб всі заявки були виконані, а загальна вартість всіх перевезень була мінімальною.

Поставимо цю задачу як задачу лінійного програмування. Позначимо x_{ij} - кількість одиниць вантажу, що відправляється з i-го ПВ A до j-го ПП B.

Сукупність чисел x_{ij} називатимемо «планом перевезень», а самі величини x_{ij} - «перевезеннями». Ці невід’ємні змінні повинні задовольняти наступним умовам:

1. Сумарна кількість вантажу, що направляється з кожного ПВ до усіх ПП, повинна дорівнювати запасу вантажу в даному пункті.

2. Сумарна кількість вантажу, що доставляється до кожного ПП із усіх ПВ, повинна дорівнювати заявці, поданої даним пунктом.

3. Сумарна вартість всіх перевезень, тобто сума величин x_{ij}, помножених на відповідні вартості c_{ij}, повинна бути мінімальною.

Отже, задача зводиться до визначення такого плану перевезень певного продукту з пунктів його виробництва до пунктів споживання $x = (x_{11}, \ldots, x_{1n}, x_{21}, \ldots, x_{2n}, \ldots, x_{i1}, \ldots, x_{in}, \ldots, x_{m1}, \ldots, x_{mn})$, який мінімізує цільову функцію

$$f(x) = \sum_{i=1}^{m} \sum_{j=1}^{n} c_{ij} x_{ij} \qquad (6.1)$$

на множині припустимих планів

$$\sum_{j=1}^{n} x_{ij} = a_i, \ \sum_{i=1}^{m} x_{ij} = b_j, \ x_{ij} \geq 0, \ i = \overline{1,m}; \quad j = \overline{1,n} \tag{6.2}$$

за умови дотримання балансу

$$\sum_{i=1}^{m} a_i = \sum_{j=1}^{n} b_j \ . \tag{6.3}$$

Якщо умова (6.3) виконується, то задачу називають **збалансованою** або **закритою**, а інакше задачу називають **незбалансованою** або **відкритою**.

Транспортна задача є представником класу задач лінійного програмування й тому має всі якості лінійних оптимізаційних задач, але одночасно вона має й ряд додаткових корисних властивостей, які дозволили розробити спеціальні методи для її розв'язання.

Число лінійно незалежних серед рівнянь в умовах-обмеженнях транспортної задачі (ТЗ) дорівнює

$$m + n - 1,$$

отже число базисних змінних також дорівнює

$$m + n - 1.$$

Загальне число змінних x_{ij} у транспортній задачі дорівнює $m*n$, а число вільних змінних

$$k = mn - (m + n - 1) = (m - 1)(n - 1).$$

Відомо, що в задачі лінійного програмування оптимальний розв'язок досягається в одній з вершин ОДР, в опорній точці, де принаймні k змінних дорівнюють нулю. Виходить, у нашому випадку для оптимального плану принаймні $(m-1)(n-1)$ перевезень повинні дорівнювати нулю (з відповідних ПВ у відповідні ПП нічого не перевозиться).

Називатимемо будь-який план перевезень **припустимим**, якщо він задовольняє умовам-обмеженням (всі заявки задоволені, всі запаси вичерпані).

Припустимий план будемо називати **опорним**, якщо в ньому відмінні від нуля не більш за $m + n-1$ базисних перевезень, а інші $(m-1)(n-1)$ перевезень дорівнюють нулю.

План називатимемо **оптимальним**, якщо він, серед всіх припустимих планів, призводить до мінімальної сумарної вартості перевезень (L=min).

В силу особливої структури ТЗ при її розв'язанні не приходиться довго розв'язувати систему рівнянь. Всі операції із знаходження оптимального плану зводяться до маніпуляцій безпосередньо з таблицею, де в певному порядку записані умови транспортної задачі: перелік ПВ й ПП, заявки й запаси, а також вартості перевезень c_{ij}. У міру заповнення цієї таблиці в її клітках проставляють самі перевезення x_{ij}. Транспортна таблиця складається з m рядків і n стовпців. Рядки транспортної таблиці відповідають пунктам виробництва (в останній клітці кожного рядка зазначений обсяг запасу продукту a_i), а стовпці - пунктам споживання (остання клітка кожного стовпця містить значення заявки b_j). У правому верхньому куті кожної клітки ставлять вартість c_{ij} перевезення одиниці продукту з A_i до B_j, а центр клітки залишають вільним, щоб поміщати до нього саме перевезення x_{ij}. Клітки, які містять нульові перевезення ($x_{ij}=0$), називають вільними, а ненульові - зайнятими ($x_{ij}>0$).

6.2 Методи побудови опорного плану

За аналогією з іншими задачами лінійного програмування розв'язання транспортної задачі починається з побудови припустимого базисного плану. Найпростіший спосіб його знаходження ґрунтується на так званому **методі північно-західного кута**. Суть методу полягає в послідовному розподілі всіх запасів, наявних у першому, другому і т.д. пунктах виробництва, до першого, другого і т.д. пунктів споживання. Кожний крок розподілу зводиться до спроби повного вичерпання запасів у черговому пункті виробництва або до спроби повного задоволення потреб у черговому пункті споживання. На кожному кроці q величини поточних нерозподілених запасів позначають a_i^q, а поточних незадоволених потреб - b_j^q. Побудова припустимого початкового плану за методом північно-західного кута починається з лівого верхнього кута транспортної таблиці. Для чергової клітки, розташованої в рядку і і стовпці j, розглядають значення нерозподіленого запасу в i-му пункті виробництва й незадоволеної заявки в j-му пункті призначення, з них вибирають мінімальне й призначають як обсяг перевезення між даними пунктами: $x_{ij} = \min\{a_i^q, b_j^q\}$. У результаті цього значення нерозподіленого запасу й незадоволеної потреби у відповідних пунктах зменшуються:

$$a_i^{(q+1)} = a_i^q - x_{ij}; \qquad b_j^{(q+1)} = b_j^q - x_{ij} .$$

Очевидно, що на кожному кроці виконується хоча б одна з рівностей: $a_i^{(q+1)} = 0$ або $b_j^{(q+1)} = 0$. Якщо справедливо $a_i^{(q+1)} = 0$, це означає, що весь запас i-го пункту виробництва вичерпаний і необхідно перейти до розподілу запасу в пункті виробництва $i + 1$, тобто переміститися до наступної клітки униз за стовпцем. Якщо $b_j^{(q+1)} = 0$, то повністю задоволена заявка для j-го пункту, після чого виконують перехід на клітку, розташовану праворуч за рядком. Знову обрана клітка стає поточною, і для неї повторюють всі перелічені операції.

Ґрунтуючись на умові балансу запасів і заявок (6.3), неважко довести, що за кінцеве число кроків буде отриманий припустимий план. У силу тієї самої умови число кроків алгоритму не може бути більшим за $m+n$-1, тому завжди залишаться вільними (нульовими) mn-($m+n$-1) кліток. Отже, отриманий план є базисним. Не виключено, що на певному проміжному кроці поточний нерозподілений запас опиниться рівним поточній незадоволеній заявці ($a_i^q = b_j^q$). У цьому випадку перехід до наступної клітки відбувається в діагональному напрямку (одночасно змінюються поточні пункти виробництва й призначення), а це означає «втрату» одного ненульового компонента в плані або, виродженість побудованого плану.

Розглянемо приклад. З 3-х пунктів виробництва необхідно вивезти однорідний продукт в 5 пунктів споживання. Транспортні витрати, обсяг виробництва й споживання наведені в таблиці 6.1.

Таблиця 6.1 - Вихідні дані

	B_1	**B_2**	**B_3**	**B_4**	**B_5**	**Запаси**
A_1	7	5	2	8	7	125
A_2	8	9	4	6	9	60
A_3	5	1	9	2	3	115
Заявки	30	50	100	40	80	300

Зауважимо, що запаси дорівнюють заявкам. Отже, задача є збалансованою.

Визначимо кількість базисних змінних

$$m + n - 1 = 3 + 5 - 1 = 7.$$

Кількість вільних змінних

$$(n-1)*(m-1) = (5-1)*(3-1) = 8.$$

Іншими методами визначення початкового опорного плану є метод мінімальної вартості, метод подвійної переваги або метод апроксимації Фогеля.

У таблиці 6.2 показаний процес пошуку припустимого плану за методом північно-західного кута, включаючи послідовну зміну обсягу нерозподілених запасів і незадоволених потреб. Стрілки відображають траєкторію переходу по клітках транспортної таблиці, а цифри, що знаходяться за її межами, - поточні нерозподілені залишки після призначення обсягу для чергової клітки.

Таблиця 6.2 - Визначення опорного плану за методом північно-західного кута

	B_1	**B_2**	**B_3**	**B_4**	**B_5**	**Запаси**	
A_1	30 ⇒	50 ⇒	45 ⇓			125	95 45 0
A_2			55 ⇒	5 ⇓		60	5 0
A_3				35 ⇒	80	115	80 0
Заявки	30	50	100	40	80	300	
	0	0	55				
			0	35			
				0	0		

Знайдений опорний план x = (30, 50, 45, 0, 0, ...). Значення цільової функції при цьому плані перевезень

$$L = 30*7+50*5+45*2+55*4+5*6+35*2+80*3 = 1110.$$

Особливістю припустимого плану побудованого за методом північно-західного кута є те, що цільова функція на ньому приймає значення, як правило, не оптимальне. Це відбувається тому, що при його побудові ніяк не враховуються значення c_{ij}. У зв'язку з цим на практиці для одержання вихідного плану використовують інший спосіб - метод мінімального елемента, у якому при розподілі обсягів перевезень в першу чергу займають клітки з найменшими цінами.

Метод мінімальної вартості полягає в тому, що в таблиці вартостей вибирають найменшу, і в цій клітці записують найменше з чисел (a_i, b_j), таблиця 6.3.

Таблиця 6.3 - Визначення опорного плану за методом мінімальної вартості

	B_1	**B_2**	**B_3**	**B_4**	**B_5**	**Запаси**
A_1	7 **25**	5	2 **100**	8	7	125
A_2	8 **5**	9	4	6	9 **55**	60
A_3	5	1 **50**	9	2 **40**	3 **25**	115
Потреби	30	50	100	40	80	300

Значення цільової функції при цьому плані перевезень

$$L = 25*7+100*2+5*8+55*9+50*1+40*2+25*3 = 1115.$$

Метод подвійної переваги показаний у таблиці 6.4. Він полягає в тому, що в кожному стовпці позначають знаком *V* клітку з найменшою вартістю, потім те саме роблять у кожному рядку. У клітки з *W* заносять найбільші обсяги перевезень. Потім розподіляють перевезення по клітках, позначених *V*. У частині таблиці, що залишилася, перевезення розподіляють за методом найменшої вартості.

Таблиця 6.4 - Визначення опорного плану за методом подвійної переваги

	B_1	**B_2**	**B_3**	**B_4**	**B_5**	**Запаси**
A_1	7 **25**	5	2 **100** W	8	7	125
A_2	8 **5**	9	4 V	6	9 **55**	60
A_3	5 V	1 **50** W	9	2 **40** V	3 **25** V	115
Заявки	30	50	100	40	80	300

Отриманий опорний план збігається з планом за методом мінімальної вартості, L = 1115.

Приймати як опорний треба той план, для якого транспортні витрати опинилися найменшими. Отже, за опорний треба прийняти план, отриманий за методом північно-західного кута.

Цей план є припустимим, тому що суми за рядками дорівнюють запасам, а суми за стовпцями дорівнюють заявкам.

Отриманий план є опорним, тому що число ненульових перевезень дорівнює $m+n$-1=3+5-1=7, а число нульових перевезень дорівнює (n-1)*(m-1)= =(5-1)*(3-1) = 8.

План можна покращити (табл. 6.2), якщо зменшити перевезення в дорогій клітці, наприклад (1.1), і збільшити в дешевій (3.1). Щоб план залишався опорним, необхідно одну з вільних кліток зробити базисною, а одну з базисних - вільною.

6.3 Метод потенціалів

Одним з методів розв'язання транспортної задачі є метод, що одержав назву **методу потенціалів**. Уперше він був запропонований в 1949 р. Л. В. Канторовичем і М. К. Гавуріним. Пізніше на базі загальних ідей лінійного програмування аналогічний метод був запропонований Дж. Данцигом.

Так само як транспортна задача є окремим випадком задачі ЛП, так і метод потенціалів, загалом кажучи, може трактуватися як різновид симплексних процедур. Він являє собою ітеративний процес, на кожному кроці якого розглядають певний поточний базисний план, перевіряють його оптимальність, і при необхідності здійснюють, перехід до «кращого» базисного плану.

Алгоритм методу потенціалів починається з вибору певного припустимого базисного плану. Якщо початковий опорний план має $\boldsymbol{m+n-1}$ додатних перевезень, то його називають **невиродженим**. Якщо опорний план має менше за $\boldsymbol{m+n-1}$ додатних перевезень, то його називають **виродженим**.

У цьому початковому опорному плані кожному пункту ставлять у відповідність певне число, називане його **попереднім потенціалом**. Якщо даний план невироджений (число ненульових базисних кліток дорівнює $m+n-1$), то за ним можна так визначити потенціали u_i і v_j, щоб для кожної базисної клітки (тобто для тієї, в якій $x_{ij}>0$) виконувалася умова

$$v_j - u_i = c_{ij}, \quad \text{якщо } x_{ij} > 0. \tag{6.4}$$

Теорема 6.1. Якщо план ТЗ є оптимальним, то йому відповідає система з $m+n$ чисел, що задовольняють умовам:

$$u_i + v_j = c_{ij} \quad \text{для} \quad x_{ij} > 0,$$

$$u_i + v_j \le c_{ij} \quad \text{для} \quad x_{ij} = 0, \qquad i = \overline{1,m}; \quad j = \overline{1,n}\,.$$

де u_i і v_j - потенціали постачальників і споживачів відповідно.

Отже, якщо хоча б для однієї вільної клітки

$$\Delta_{ij} = u_i + v_j - c_{ij} > 0, \tag{6.5}$$

план не є оптимальним і вимагає поліпшення.

Оскільки система (6.4) містить $m+n-1$ рівнянь й $m+n$ невідомих, то один з потенціалів можна задати довільно (наприклад, дорівняти v_j або u_i, до нуля). Після цього інші невідомі u_i і v_j - визначаються однозначно.

Розглянемо процес визначення потенціалів поточного плану транспортної задачі на прикладі. У таблиці 6.5 переписані умови задачі з таблиці 6.1 і її припустимий базисний план, побудований за методом північно-західного кута з таблиці 6.2.

Потенціал першого пункту виробництва приймаємо рівним нулю ($u_1=0$). Тепер, знаючи його, можна визначити потенціали для всіх пунктів призначення, пов'язаних з першим пунктом виробництва ненульовими перевезеннями. У цьому випадку їх три (це перший, другий і третій пункти), отримаємо:

$$v_1 = c_{11}-u_1= 7 - 0 = 7,\ v_2 = c_{12}-u_1=5-0 = 5,\ v_3 = c_{13}-u_1=2-0 = 2.$$

Маючи v_3 і з огляду на те, що в другому рядку таблиці існують ненульові компоненти x_{23} і x_{24}, можна визначити $u_2 = c_{23}- v_3 =4 - 2 =2$, $v_4= c_{24}- u_2 =6 - 2 = 4$, після чого з'являється можливість розрахувати $u_3=c_{34}-v_4 =2-4 = -2$ і, нарешті,

$v_5 = c_{35}$-u_3 =3 - (-2) = 5. В результаті отримали повну систему потенціалів, показану в таблиці 6.5.

Таблиця 6.5 - Визначення потенціалів для початкового опорного плану

		v_1=7	v_2=5	v_3=2	v_4=4	v_5=5	
		B_1	**B_2**	**B_3**	**B_4**	**B_5**	**Запаси**
u_1=0	**A_1**	7 30	5 50	2 45	8	7	125
u_2=2	**A_2**	8	9	4 55	6 5	9	60
u_3=-2	**A_3**	5	1	9	2 35	3 80	115
	Заявки	30	50	100	40	80	300

Для вільних кліток транспортної таблиці обчислюють величини $\Delta_{ij}=v_j+u_i- c_{ij}$. У таблиці 6.6 вони виписані для всіх небазисних кліток під цінами.

Таблиця 6.6 - Перевірка оптимальності поточного плану (обчислення оцінок Δ_{ij})

		v_1=7	v_2=5	v_3=2	v_4=4	v_5=5	
		B_1	**B_2**	**B_3**	**B_4**	**B_5**	**Запаси**
u_1=0	**A_1**	7 30	5 50	2 45	8 0	7 -2	125
u_2=2	**A_2**	8 1	9 -2	4 55	6 5	9 -2	60
u_3=-2	**A_3**	5 0	1 2	9 -9	2 35	3 80	115
	Заявки	30	50	100	40	80	300

Відповідно до теореми 6.1, якщо всі $\Delta_{ij}\leq 0$, план є оптимальним, у протилежному випадку, якщо існує хоча б одна клітка, для якої $\Delta_{ij} > 0$, його можна поліпшити. Процес «поліпшення» плану полягає у визначенні клітки, що вводиться, і клітки, що виводиться. У цьому простежується змістовна аналогія методу з відповідними пунктами симплекс-процедур.

Кандидатом на введення може бути будь-яка клітка, у якій $\Delta_{ij} > 0$, оскільки після введення її до базису буде забезпечена рівність $v_j + u_i = c_{ij}$. Для визначеності рекомендується брати ту клітку, у якій оцінка Δ_{ij} максимальна. У розглянутому нами прикладі це буде клітка (3, 2).

Виведена клітка визначається за допомогою так званого ланцюжка перетворення плану, що описує характер перерозподілу вантажних потоків. У відповідності із властивостями транспортної задачі для невиродженого базисного плану в поточній таблиці можна утворити замкнутий ланцюжок, що складається з вертикальних і горизонтальних ланок, однією з вершин якої є обрана вільна клітка, а інші – зайняті клітки. У таблиці 6.7 показаний ланцюжок перетворення поточного плану щодо клітки, яка вводиться до нього, (3, 2).

Таблиця 6.7 - Перетворення поточного плану

		$v_1=7$	$v_2=5$	$v_3=2$	$v_4=4$	$v_5=5$	
		B_1	**B_2**	**B_3**	**B_4**	**B_5**	**Запаси**
$u_1=0$	**A_1**	7 30	5 50 -Θ	2 45 +Θ	8 4	7 5	125
$u_2=2$	**A_2**	8 5	9 3	4 55 -Θ	6 5 +Θ	9 3	60
$u_3=-2$	**A_3**	5 9	1 +Θ 7	9 4	2 35 -Θ	3 80	115
	Заявки	30	50	100	40	80	300

Логіка алгоритму побудови ланцюжка досить проста: «вийшовши» з клітки (3, 2) у горизонтальному напрямку, ми повинні «зупинитися» у тій зайнятій клітці плану, з якої зможемо рухатися далі за вертикаллю. У даному прикладі цій вимозі задовольняють як клітка (3, 4), так і клітка (3, 5). Проте ланцюжок від (3, 5) не може бути продовжений далі, у той час як рухаючись від (3, 4) за вертикаллю до (2, 4) і далі до (2.3), ми вертаємося через клітки (1, 3) і (1, 2) до вихідної клітки (3, 2) і утворюємо замкнений цикл.

У побудованому ланцюжку, починаючи з клітки, що вводиться (яка вважається першою), позначають вершини: непарні - «+Θ», а парні «-Θ». Знаком «+» позначають ті клітки, у яких обсяги перевезень повинні збільшитися (такою, зокрема, є клітка, що вводиться до плану, оскільки вона повинна стати базисною). Знаком «-» - ті клітки, у яких перевезення зменшуються з метою збереження балансу. Серед множини кліток, позначених знаком «-», обирають клітку з найменшим значенням x_{ij}. Вона і стає кандидатом на вивід, тому що зменшення обсягу перевезень на більшу величину може призвести до від'ємних значень x_{ij} в інших «мінусових» клітках. Потім провадиться перерахування плану за ланцюжком: до обсягів перевезень у клітках, позначених знаком «+», додається обсяг Θ, а з обсягів кліток, позначених знаком «-», він віднімається. У результаті вводу однієї клітки й виводу іншої утворюється новий базисний план, для якого на наступній ітерації описані вище дії повторюють.

У нашому прикладі знаком «-» позначені клітки (3, 4), (2, 3) і (1, 2), причому x_{34}=35, x_{23} =55, x_{12} =50. Обчисливши значення $\Theta = \min\{x_{34}, x_{23}, x_{12}\}$ =35, здійснюємо перетворення і переходимо до наступного базисного плану, показаного в таблиці 6.8.

Таблиця 6.8 - Оцінка оптимальності наступного базисного плану

		$v_1=7$	$v_2=5$	$v_3=2$	$v_4=4$	$v_5=7$	
		B_1	**B_2**	**B_3**	**B_4**	**B_5**	**Запаси**
$u_1=0$	**A_1**	7 30	5 15	2 80	8 -4	7 0	125
$u_2=2$	**A_2**	8 1	9 -2	4 20	6 40	9 0	60
$u_3=-4$	**A_3**	5 -2	1 35	9 -11	2 -2	3 80	115
	Заявки	30	50	100	40	80	300

Для знов отриманого плану повторюють дії стандартної ітерації: розраховують потенціали й оцінки для небазисних кліток транспортної таблиці. Як можна бачити, план у таблиці 6.8 також не є оптимальним (у клітці (2, 1) Δ_{21} =1>0), тому знову будуємо ланцюжок перетворення плану й переходимо до наступного базисного плану за ланцюжком в таблиці 6.9.

Таблиця 6.9 - Перетворення поточного базисного плану

		$v_1=7$	$v_2=5$	$v_3=2$	$v_4=4$	$v_5=7$	
		B₁	**B₂**	**B₃**	**B₄**	**B₅**	**Запаси**
$u_1=0$	**A₁**	7 30 $-\Theta$	5 15	2 80 $+\Theta$	8 -4	7 0	125
$u_2=2$	**A₂**	8 $+\Theta$ 1	9 -2	4 20 $-\Theta$	6 40	9 0	60
$u_3=-4$	**A₃**	5 -2	1 35	9 -11	2 -2	3 80	115
	Заявки	30	50	100	40	80	300

Визначивши Θ=20, одержимо показаний в таблиці 6.10 оптимальний базисний план.

Таблиця 6.10 - Оптимальний базисний план

		$v_1=7$	$v_2=5$	$v_3=2$	$v_4=5$	$v_5=7$	
		B₁	**B₂**	**B₃**	**B₄**	**B₅**	**Запаси**
$u_1=0$	**A₁**	7 10	5 15	2 100	8 -3	7 0	125
$u_2=1$	**A₂**	8 20	9 -3	4 -1	6 40	9 -1	60
$u_3=-4$	**A₃**	5 -2	1 35	9 -11	2 -1	3 80	115
	Заявки	30	50	100	40	80	300

З транспортної таблиці 6.10 видно, що отриманий план оптимальний, тому що всі оцінки для небазисних кліток $\Delta_{ij} \le 0$, тобто $u_i + v_j$ не перевищують відповідних цін c_{ij}. За цим планом обчислюють оптимальне (найменше) значення сумарних витрат на перевезення

$$L^*=10x_7+15x_5+100x_2+20x_8+40x_6+35x_1+80x_3=1020.$$

6.4 Виродження плану

Зупинимося на ситуації виникнення виродженого плану. Якщо задача вироджена, то на якомусь етапі розв’язання може виявитися, що таблиця містить менше за $m+n$-1 заповнених кліток. Це, зокрема, може відбутися, якщо те саме мінімальне значення буде досягнуто відразу на кількох клітках, позначених знаком «-». Для подолання виродженості в транспортній задачі поточний план доповнюють необхідною кількістю нульових кліток (фіктивними перевезеннями) так, щоб вони дозволяли розрахувати повну систему потенціалів, і далі діяти відповідно до правил описаного вище алгоритму. Тобто, якщо не вистачає k заповнених кліток, то їх вважають фіктивно заповненими. Фактично такий прийом є аналогом методу збурень для транспортної задачі як окремого випадку ЗЛП. До такого висновку легко прийти, якщо покласти, що фіктивні клітки, що додаються, містять певний малий обсяг ε.

Розглянемо приклад. З трьох кар'єрів до чотирьох склозаводів возять глину. Вартості перевезень, потужності кар'єрів і потреби заводів наведені в таблиці 6.11.

Таблиця 6.11 - Вихідні дані

	B_1	B_2	B_3	B_4	**Запаси**
A_1	3	9	7	4	50
A_2	6	8	10	6	65
A_3	5	4	7	6	50
Потреби	50	20	65	30	165

Опорний план визначимо за методом найменшої вартості.

Таблиця 6.12 - Початковий опорний план

		$v_1=3$	$v_2=5$	$v_3=8$	$v_4=4$	
		B_1	B_2	B_3	B_4	**Запаси**
$u_1=0$	A_1	3 50	9 -4	$+\Theta$ 7 1	$-\Theta$ 4 [0]	50
$u_2=2$	A_2	6 -1	8 -1	$-\Theta$ 10 35	$+\Theta$ 6 30	65
$u_3=-1$	A_3	5 -3	4 20	7 30	6 -3	50
	Потреби	50	20	65	30	165

У таблиці 6.12 заповнених кліток 5, а необхідно $m+n-1 = 3+4-1 = 6$. Потрібна одна фіктивно заповнена клітка. Вважатимемо клітку (1, 4) заповненою. Перевіримо план на оптимальність.

План не оптимальний, оскільки $\Delta_{13} = 1$. $\Theta = \min(35,0) = 0$. План залишається таким самим, але фіктивно заповненою буде клітка (1, 3). Перевіримо його на оптимальність.

Таблиця 6.13 - Перевірка базисного плану на оптимальність

		$v_1=3$	$v_2=4$	$v_3=7$	$v_4=3$	
		B_1	B_2	B_3	B_4	**Запаси**
$u_1=0$	A_1	3 50	9 -5	7 0	4 -1	50
$u_2=3$	A_2	6 0	8 -1	10 35	6 30	65
$u_3=0$	A_3	5 -2	4 20	7 30	6 -3	50
	Потреби	50	20	65	30	165

Всі $\Delta_{ij} \leq 0$, отже план є оптимальним. Вартість перевезень при такому плані становить $L = 50*3+20*4+0*7+35*10+30*7+30*6 = 970$.

6.5 Транспортна задача за критерієм часу

Транспортну задачу, окрім використання її з метою мінімізації витрат при транспортуванні вантажів, застосовують для оптимізації цілого ряду економічних процесів.

Природним застосуванням транспортної задачі є **задача мінімізації часу транспортування**. Наприклад, на підприємстві є задана кількість виробничих ліній, що працюють із заданою продуктивністю. Є також задана кількість складів продукції. Відомий час транспортування від кожної лінії до кожного складу. Задача зводиться до вибору таких маршрутів, при використанні яких час, затрачуваний на транспортування, буде мінімальним.

Іншою задачею є вибір оптимального варіанта з використання наявного устаткування. Задача полягає у виборі оптимального плану виробництва виробів при мінімізації часу, необхідного на їх виготовлення. Цю задачу формулюють в такий спосіб. Є задана кількість видів устаткування, на кожному з яких виготовляють певну задану кількість виробів. Відома продуктивність для кожного типу устаткування й кожного виробу. Необхідно скласти план, при якому підприємство затратить мінімальну кількість часу на виготовлення всіх виробів.

Розглянемо приклад. Визначити мінімальний час виготовлення трьох видів виробів на трьох типах устаткування. Продуктивність устаткування (a_i) і план виробництва виробів (b_j) наведені в таблиці 6.14.

Таблиця 6.14 - Вихідні дані

	B_1	**B_2**	**B_3**	**План**
A_1	15	7	8	240
A_2	9	4	11	80
A_3	6	3	7	180
Продуктивність	200	160	140	500

Задача є збалансованою. Знайдемо початковий базисний план за методом мінімальної вартості. Найменша продуктивність, що дорівнює 3, записана в клітці (3, 2). Потужність устаткування третього типу дорівнює 180, а необхідна кількість виробів другого типу дорівнює 160, тобто min{180,160}=160. Тому в клітку (3, 2) вписуємо число 160. Другий стовпець випадає з подальшого розгляду, оскільки продуктивність устаткування другого типу вичерпана. Аналогічно заповнюємо інші клітки таблиці 6.15.

Таблиця 6.15 - Визначення початкового опорного плану

		$v_1=15$	$v_2=12$	$v_3=8$	
		B_1	**B_2**	**B_3**	**План**
$u_1=0$	**A_1**	15 -Θ 100	7 +Θ 5	8 140	240
$u_2=-6$	**A_2**	9 80	4 2	11 -9	80
$u_3=-9$	**A_3**	6 +Θ 20	3 -Θ 160	7 -8	180
	Продуктивність	200	160	140	500

Відповідно до отриманого плану час виготовлення всіх виробів становить L = 100*15+140*8+80*9+20*6+160*3 = 3940. Визначимо потенціали u_i і v_j для зайнятих кліток: u_1=0; v_1 =15; v_3 =8; u_2=9-15=-6; u_3=6-15=-9; v_2 =3-(-9) = 12. Обчислимо оцінки для вільних кліток: Δ_{12}=(12+0)-7=5; Δ_{22}=(12-6)-4=2; Δ_{23} = (-6+8)-11=-9; Δ_{33}=(-9+8)-7=-8.

Оскільки оцінки вільних кліток (1, 2) і (2, 2) додатні, план не оптимальний. Перейдемо до нового базисного плану, побудувавши цикл для клітки (1, 2), Θ= min{100, 160}=100. Новий базисний план наведений у таблиці 6.16. З клітки (1,1) вироби в кількості 100 од. переносимо до клітки (1,2). З клітки (3,2) вироби в кількості 100 од. переносимо до клітки (3,1).

Таблиця 6.16 - Визначення наступного базисного плану

		$v_1=10$	$v_2=7$	$v_3=8$	
		B_1	**B_2**	**B_3**	**План**
$u_1=0$	**A_1**	15 -5	7 100	8 140	240
$u_2=-1$	**A_2**	-Θ 9 80	+Θ 4 2	11 -4	80
$u_3=-4$	**A_3**	+Θ 6 120	-Θ 3 60	7 -3	180
	Продуктивність	200	160	140	500

Відповідно до отриманого плану час виготовлення всіх виробів складе L=100*7+140*8+80*9+120*6+60*3 = 3440. Перевірка плану на оптимальність показує, що він не оптимальний, тому що в клітці (2,2) Δ_{22} додатна. Побудуємо цикл для цієї клітки й перетворимо план при Θ=60. Новий базисний план наведений у таблиці 6.17.

Таблиця 6.17 - Визначення наступного базисного плану

		$v_1=12$	$v_2=7$	$v_3=8$	
		B_1	**B_2**	**B_3**	**План**
$u_1=0$	**A_1**	15 -3	7 100	8 140	240
$u_2=-3$	**A_2**	9 20	4 60	11 -6	80
$u_3=-6$	**A_3**	6 180	3 -2	7 -5	180
	Продуктивність	200	160	140	500

Перевірка показує оптимальність плану. Відповідно до отриманого плану час виготовлення всіх виробів складе L = 100*7+140*8+20*9+60*4+180*6 = =3320, і він є мінімальним.

Контрольні запитання

1. Які специфічні властивості дозволяють виділити транспортні задачі в окремий клас з множини задач лінійного програмування?
2. Опишіть методи побудови припустимого плану транспортної задачі.
3. Скільки ненульових елементів повинен містити невироджений базисний план транспортної задачі?

4. Сформулюйте критерій оптимальності для припустимого плану транспортної задачі.

5. Поясніть, на чому заснований метод потенціалів?

6. З чого випливає критерій оптимальності припустимого плану транспортної задачі?

7. Перелічіть основні етапи методу потенціалів.

8. Які умови повинні бути дотримані при побудові ланцюжка перетворення плану за методом потенціалів?

9. Як подолати виникнення ситуації виродженості поточного плану в транспортній задачі?

ТЕМА 7
ЦІЛОЧИСЛОВІ ЗАДАЧІ ЛІНІЙНОГО ПРОГРАМУВАННЯ. ОСНОВНІ МЕТОДИ ЇХ РОЗВ'ЯЗАННЯ І АНАЛІЗУ

7.1 Типи задач дискретного програмування

Цілочислове лінійне програмування орієнтоване на розв'язання задач лінійного програмування, в яких всі або деякі змінні повинні приймати цілочислові (або дискретні) значення. Багато економічних задач характеризуються тим, що обсяги керованих ресурсів можуть приймати тільки цілі значення. До цілочислового програмування належать також задачі, в яких змінні можуть приймати тільки два значення – 0 або 1 (булеві або бінарні змінні). До задач цілочислового програмування належать задачі про призначення, про найкоротший шлях та інші. У загальному вигляді задачу цілочислового програмування можна сформулювати як задачу знаходження максимуму (або мінімуму) цільової функції $L(x_1, x_2, ..., x_n)$ на множині D, зумовленій системою обмежень. Загальну задачу цілочислового програмування формулюють в такий спосіб.

Знайти найбільше або найменше значення лінійної функції

$$L = \sum_{j=1}^{n} c_j x_j \to extr \tag{7.1}$$

за умови

$$\sum_{j=1}^{n} a_{ij} x_j = b_i; \quad i = \overline{1,m}, \tag{7.2}$$

$$x_j \geq 0, \quad x_j - цілі, \quad j = \overline{1,n}.$$

Умову $x_j - цілі$ називають умовою дискретності. Особливе місце серед дискретних задач займає цілочислова задача лінійного програмування в канонічній формі (ЦКЗЛП). У певних ситуаціях вимога «цілочислове» може бути накладена лише на певні змінні x_j, що кардинально не змінює характеру задачі.

Принципова складність, викликувана наявністю умов цілочисленості в системі обмежень оптимізаційної задачі, полягає в тому, що в значній кількості випадків неможливо замінити дискретну задачу її безперервним аналогом і, знайшовши відповідний розв’язок, округлити його компоненти до найближчих цілих значень. На рисунку 7.1 показано, що при округленні оптимального плану x^* звичайної задачі ЛП до цілих значень виходить точка B, яка не належить області припустимих планів. Якщо навіть оптимальний план безперервної задачі, округлений до цілих значень компонент, виявиться припустимим, то цільова функція може поводитися так, що її значення буде на ньому істотно «гірше», ніж на оптимальному плані цілочислової задачі.

Перелічені проблеми визначили необхідність розробки спеціальних методів розв’язання дискретних і цілочислових задач.

У літературі, як правило, виділяють наступні класи дискретних оптимізаційних задач:

- задача з неділимостями;
- екстремальні комбінаторні задачі;
- задачі з розривними цільовими функціями;
- задачі на незв'язних і неопуклих областях та ін.

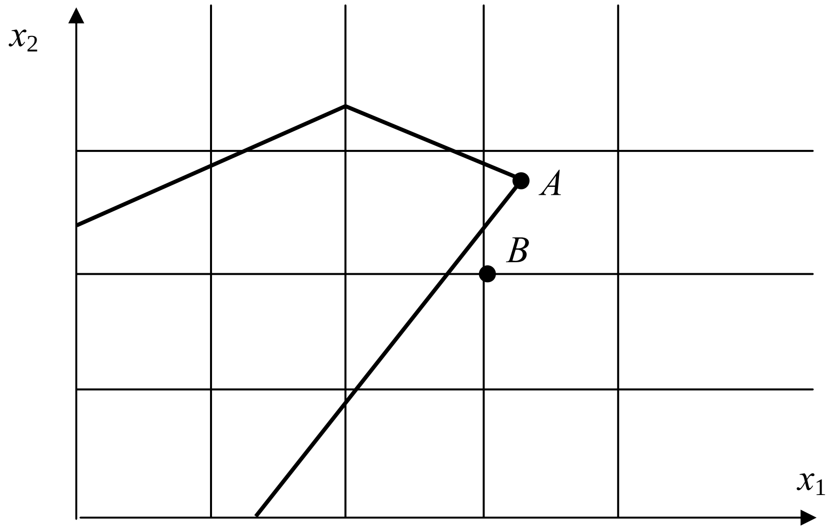


Рисунок 7.1 – Розв’язання цілочислової задачі

Задача з неділимостями. У переважній більшості випадків наявність умов неділимості визначається фізичними властивостями об'єктів, наприклад, вони можуть з'явитися як додаткові обмеження в задачі виробничого планування, якщо в ній здійснюється керування випуском великої одиничної продукції. Класичним представником задач даного класу є так звана задача про ранець. Вона полягає в тому, що солдат (або турист), що збирається в похід, може нести вантаж вагою не більш за W кг. Цей вантаж може складатися з набору предметів n типів, кожний предмет j-го типу важить w кг і характеризується певною «корисністю». Треба визначити, скільки предметів кожного виду потрібно покласти в ранець, щоб його сумарна корисність була максимальною. До такого формулювання можуть бути зведені багато економічних задач. У літературі ця задача також відома як задача про завантаження судна.

Комбінаторні задачі. До даного класу належать задачі з оптимізації функції, заданої на кінцевій множині, елементами якої служать вибірки з n об'єктів.

Класичним представником такої задачі є задача про комівояжера. Вона полягає в складанні маршруту відвідування торговим агентом, що знаходиться в певному початковому пункті, *n* інших міст за умови, що задано матрицю вартостей переїздів з одного міста до іншого. Причому припустимим є маршрут, який передбачає однократне відвідування всіх міст і повернення у вихідний пункт. Найкращий маршрут повинен мінімізувати сумарну вартість переїздів. Кожний припустимий маршрут можна ототожнити з перестановкою *n* чисел. Задача комівояжера має велику кількість змістових аналогів. Зокрема, задача розробки графіка переналагодження устаткування, що може випускати різні типи виробів, але вимагає певних витрат (часових або матеріальних) при переході з одного технологічного режиму на інший.

Задача з розривними цільовими функціями. Багато економічних систем характеризуються наявністю так званих постійних витрат, які мають місце незалежно від обсягу виробництва. Врахування у моделях цих і подібних факторів призводить до появи в них цільових функцій, які не є безперервними. Як приклад можна навести транспортну задачу з фіксованими доплатами. У цьому випадку цільова функція сумарних витрат на перевезення містить «стрибкоподібні» розриви, що істотно утрудняє її мінімізацію.

Методи розв’язання задач цілочислового лінійного програмування засновані на використанні обчислювальних можливостей методів лінійного програмування. Зазвичай алгоритми цілочислового програмування включають три кроки:

1. «Ослаблення» простору припустимих розв’язків задачі шляхом відкидання вимоги цілочисленості. У результаті виходить звичайна задача лінійного програмування.

2. Визначення оптимального розв’язку задачі лінійного програмування.

3. Маючи отриманий оптимальний розв’язок, додають спеціальні обмеження, які ітераційним шляхом змінюють простір припустимих розв’язків задачі лінійного програмування так, щоб в остаточному підсумку вийшов оптимальний розв’язок, який задовольняє вимогам цілочисленості.

Розроблено два методи для побудови спеціальних обмежень:

- метод Гоморі (або метод площин, що відсікають, він був уперше запропонованим Р. Гоморі у 1957-1958 р.);
- метод віток і границь.

Помітимо також, що досить ефективний і широко застосовуваний підхід до розв’язання цілочислових задач заснований на зведенні їх до задач транспортного типу. Це пояснюється тим, що якщо в умовах транспортної задачі значення запасів і потреб є цілочисловими, то цілочисловим буде й оптимальний план.

7.2 Метод Гоморі

Сутність методу Гоморі полягає в тому, що спочатку задачу вирішують без урахування умови цілочисленості. Якщо отриманий у результаті оптимальний план x^* містить тільки цілі компоненти, задачу вирішено. У протилежному випадку до системи обмежень задачі додають нове обмеження, що має наступні властивості:

- воно повинне бути лінійним;

- повинне відсікати знайдений нецілочисловий оптимальний план x^*;

- не повинне відсікати жодного цілочислового плану.

Додаткове обмеження, що має зазначені властивості, називають **правильним відсіканням**. Далі задачу вирішують з урахуванням нового обмеження. Після цього, якщо буде потреба, додають ще одне обмеження і т. д.

Геометрично додавання кожного лінійного обмеження відповідає проведенню прямої (гіперплощини), що відсікає від багатокутника (багатогранника) розв'язків певну його частину разом з оптимальною точкою з нецілими координатами, але не зачіпає ні однієї з цілих точок цього багатогранника. В результаті новий багатогранник розв'язків містить всі цілі точки, утримувані в первісному багатограннику розв'язків, і відповідно отриманий при цьому багатограннику оптимальний розв'язок буде цілочисловим.

Один з алгоритмів розв'язання задачі лінійного цілочислового програмування, запропонований Гоморі, заснований на симплексному методі й використовує досить простий спосіб побудови правильного відсікання.

Нехай задача лінійного програмування має кінцевий оптимум, і на останньому кроці її розв'язання за симплексним методом отримані наступні рівняння, у яких базисні змінні x_1, x_2, …, x_m виражені через вільні змінні x_{m+1}, x_{m+2}, …, x_n оптимального розв'язку

$$\begin{aligned} x_1 &= b_1 - a_{1\,m+1}x_{m+1} - \ldots - a_{1n}x_n, \\ x_2 &= b_2 - a_{2\,m+1}x_{m+1} - \ldots - a_{2n}x_n, \\ &\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ x_m &= b_m - a_{m\,m+1}x_{m+1} - \ldots - a_{mn}x_n, \end{aligned} \qquad (7.3)$$

де оптимальним розв'язком задачі є $x^* = (b_1, b_2, \ldots, b_m, 0, 0, \ldots, 0)$, в якому, наприклад, b_i – нецілий компонент.

Нагадаємо, що цілою частиною числа b називають найбільше ціле число $[b]$, що не перевищує b, а дробовою частиною числа b - число $\{b\}$, що дорівнює різниці між цим числом і його цілою частиною, тобто $\{b\} = b - [b]$.

Сформувавши за i-м рівнянням системи (7.3) додаткове обмеження

$$\{b_i\} - \{a_{i\,m+1}\}x_{m+1} - \ldots - \{a_{in}\}x_n \le 0, \qquad (7.4)$$

можна довести, що воно має всі властивості правильного відсікання.

Для розв'язання задачі цілочислового програмування за методом Гоморі використовують наступний алгоритм:

1. Вирішують задачу за симплексним методом без урахування умови цілочисленості. Якщо всі компоненти оптимального плану цілі, він є оптимальним і для задачі цілочислового програмування. Якщо перша задача є нерозв'язною (тобто не має кінцевого оптимуму або умови її суперечливі), то і друга задача також нерозв'язна.

2. Якщо серед компонентів оптимального розв'язку є нецілі, то вибирають компоненту з найбільшою цілою частиною і за відповідним їй рівнянням системи (7.3) формують правильне відсікання (7.4).

3. Отриману нерівність перетворюють на рівносильне рівняння шляхом введення додаткової невід’ємної цілочислової змінної

$$\{b_i\} - \{a_{i\,m+1}\}x_{m+1} - \ldots - \{a_{in}\}x_n + x_{n+1} \qquad (7.5)$$

і включають його до системи обмежень. Необхідно пам'ятати, що після включення до системи обмежень додаткового рівняння, яке відповідає правильному відсіканню, завжди буде виходити неприпустимий базисний розв’язок. Для одержання припустимого базисного розв’язку потрібно перевести в базисні змінні одну з вільних змінних.

4. Отриману розширену задачу вирішують за симплексним методом. Якщо знайдений оптимальний план буде цілочисловим, то задачу цілочислового програмування вирішено. У протилежному випадку треба повернутися до п.2 алгоритму.

Якщо задачу розв'язано в цілих числах, то після кінцевого числа кроків оптимальний цілочисловий план буде знайденим.

Якщо в процесі розв’язання з'явиться рівняння (що виражає базисну змінну через вільні) з нецілим вільним членом і цілими іншими коефіцієнтами, то відповідне рівняння не має розв’язання в цілих числах. У цьому випадку й задача не має цілочислового оптимального розв’язку.

Розглянемо особливості застосування методу Гоморі на прикладі. Нехай є задача з наступними умовами:

$$L = 3x_1 + x_2 + 2x_3 \to max,$$
$$x_1 + 2x_2 + 4x_3 = 9,$$
$$3x_1 + x_2 + 5x_3 = 10,$$
$$x_j \ge 0, \quad x_j - цілі, \quad j = \overline{1,3}.$$

Використовуючи звичайний симплекс-алгоритм, вирішуємо безперервний аналог вихідної задачі, в якої ігноруються умови цілочисленості. Як вихідний базис можна взяти перший і другий стовпці. На його основі заповнюємо симплекс-таблицю 7.1.

Таблиця 7.1 – Вихідні дані

Базис	$C_{баз}$	C_j	3	1	2
		B	A_1	A_2	A_3
A_1	3	11/5	1	0	6/5
A_2	1	17/5	0	1	7/5
	L_j	10	3	1	5
	Δ_j		0	0	3

Як видно з рядка оцінок, цей базис є оптимальним, однак відповідний йому план $x^* = \left(\frac{11}{5}; \frac{17}{5}; 0\right)$ не є цілочисловим, тому вибираємо рядок, що містить перший нецілий елемент, і відповідно до формули (7.4) будуємо відсікаюче обмеження:

$$\left\{\frac{11}{5}\right\}-\left\{\frac{6}{5}\right\}x_3 \le 0 .$$

Увівши додаткову цілочислову змінну $x_4 \ge 0$, одержимо рівносильне нерівності рівняння

$$-\frac{1}{5}x_3+x_4=-\frac{1}{5},$$

яке необхідно включити до системи обмежень. Введемо його до симплекс-таблиці 7.2.

Таблиця 7.2 – Введення нового рівняння

Базис	$C_{баз}$	C_j	3	1	2	0
		B	A_1	A_2	A_3	A_4
A_1	3	11/5	1	0	6/5	0
A_2	1	17/5	0	1	7/5	0
A_4	0	-1/5	0	0	-1/5	1
	L_j	10	3	1	5	0
	Δ_j		0	0	3	0

Врахуємо, що отриманий базисний розв’язок є неприпустимим (компонента b_3 від’ємна). Для одержання припустимого базисного розв’язку переведемо в базисні змінні вільну змінну x_3.

Таблиця 7.3 - Переведення в базисні змінної x_3

Базис	$C_{баз}$	C_j	3	1	2	0
		B	A_1	A_2	A_3	A_4
A_1	3	1	1	0	0	6
A_2	1	2	0	1	0	7
A_3	2	1	0	0	1	-5
	L_j	7	3	1	2	15
	Δ_j		0	0	0	15

Отриманий план $x^*=(1, 2, 1, 0)$ є не тільки оптимальним (всі $b_i>0$), але й повністю складається з цілочислових компонент, тобто розв’язок задачі знайдено, $\max L = 7$.

7.3 Метод віток і границь

Метод віток і границь - один з комбінаторних методів. Його суть полягає у впорядкованому переборі варіантів і розгляді тих з них, які за певними ознаками є перспективними. Вперше цей метод запропонували в 1960 р. Ленг і Дойг, а його «друге народження» відбулося в 1963 р. у зв'язку з виходом роботи Літтла, Мурті, Суїні та Керел, присвяченої розв’язанню задачі про комівояжера.

Метод віток і границь полягає в тому, що множину припустимих розв’язків (планів) за певним способом розбивають на підмножини, кожну з яких за тим самим способом знову розбивають на підмножини. Процес триває доти, поки не буде отриманий оптимальний цілочисловий розв'язок вихідної задачі.

Якщо задача максимізації лінійної функції L вирішена за симплексним методом без урахування цілочисленості змінних, то стають відомими нижня й верхня границя для кожної цілочислової змінної x_i: $[x_i]<x_i<[x_i]+1$ і нижня границя лінійної функції L_0, тобто при будь-якому плані x $L(x)\geq L_0$. Тоді з області припустимих значень, наприклад, змінної x_r виключають область $[x_r]<x_r<[x_r]+1$, де $[x_r]$ – ціла частина числа x_r. В результаті це дозволяє сформулювати дві задачі, що відрізняються тим, що до одної з них додано обмеження $x_r \leq [x_r]$, а до іншої – обмеження $x_r \geq [x_r]+1$.

Сформульовані задачі вирішують у будь-якому порядку. Залежно від отриманого розв'язку список таких задач може розширюватися або зменшуватися. Якщо в ході розв'язання будь-якої із задач отриманий нецілочисловий оптимальний план, для якого $L(x) \leq L_0$, то цю задачу виключають зі списку. Якщо $L(x) \geq L_0$, то з цієї задачі формулюють дві нові задачі.

Розглянемо приклад:

$$L = 5x_1 + 4x_2 \rightarrow max,$$
$$x_1 + x_2 \leq 5,$$
$$10x_1 + 6x_2 \leq 45,$$
$$x_j \geq 0, \quad x_j - цілі, \quad j = \overline{1,2}.$$

Вирішимо задачу шляхом відкидання умов цілочисленості. Скористаємось функцією Excel «Пошук рішення» і одержимо

x_1=3,75; x_2=1,25; L=23,75.

Виберемо одну з цілочислових змінних, значення якої в оптимальному розв'язку не є цілочисловим, наприклад, x_1. Очевидно, що область $3<x_1<4$ простору припустимих розв'язків не містить цілочислових значень змінної x_1, і отже може бути виключеною з розгляду. Це еквівалентно заміні вихідної задачі двома новими задачами. В одну з них додамо обмеження $x_1\leq 3$, одержимо

$$L = 5x_1 + 4x_2 \rightarrow max,$$
$$x_1 + x_2 \leq 5,$$
$$10x_1 + 6x_2 \leq 45,$$
$$x_1 \leq 3,$$
$$x_j \geq 0, \quad x_j - цілі, \quad j = \overline{1,2}.$$

Використовуючи функцію Excel «Пошук рішення», дістанемо оптимальний цілочисловий план

x_1=3; x_2=2; L=23.

Якість отриманого цілочислового розв'язку оцінити неможливо, оскільки розв'язок другої задачі може привести до більшого значення цільової функції. Це треба перевірити. Сформулюємо другу задачу, додавши до неї обмеження $x_1 \geq 4$:

$$L = 5x_1 + 4x_2 \to max,$$

$$x_1 + x_2 \leq 5,$$

$$10x_1 + 6x_2 \leq 45,$$

$$x_1 \geq 4,$$

$$x_j \geq 0, \quad x_j - цілі, \quad j = \overline{1,2}.$$

В оптимальному розв'язку другої задачі змінна x_2 не є цілим числом:

$x_1=4$; $x_2=0{,}83$; $L=23{,}33$.

Оскільки значення змінної $x_2=0{,}83$ не є цілим числом, другу задачу необхідно досліджувати далі, розділивши її у свою чергу ще на дві задачі і т.д. Проілюструємо хід розв'язання схемою, наведеною на рисунку 7.2.

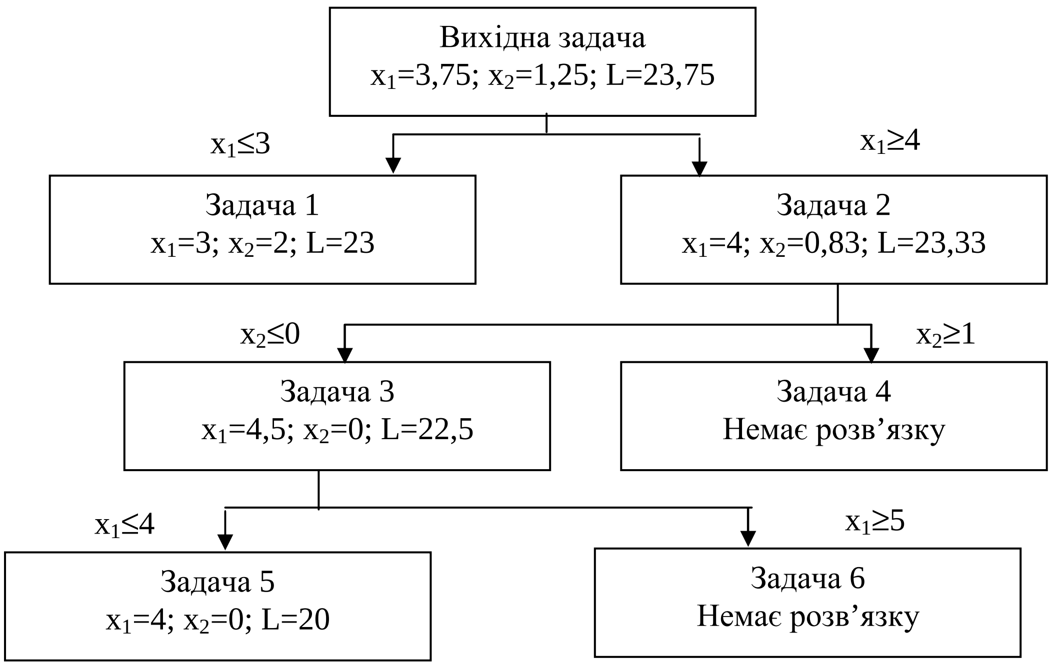


Рисунок 7.2 – Схема методу віток і границь

У ході розв'язання можна було спочатку як змінну розгалуження прийняти x_2. При цьому наступні обчислення можуть значно відрізнятися.

Очевидним недоліком алгоритму методу віток і границь при розв'язанні задач великої розмірності є необхідність перебрати занадто велику кількість варіантів, перш ніж буде знайдений оптимальний план.

Контрольні запитання

1. Які основні проблеми виникають при розв'язанні дискретних задач?
2. Сформулюйте задачу про ранець.
3. Які економіко-математичні моделі можна звести до задачі про комівояжера?
4. Наведіть приклади моделей з розривними цільовими функціями.

5. Який принцип використовують для побудови правильного відсікання в методі Гоморі?
6. Яку роль відіграє алгоритм двоїстого симплекс-методу при розв’язанні цілочислової лінійної задачі за методом Гоморі?
7. Перелічіть принципові ідеї, що лежать в основі методу віток і границь.
8. Як провадиться побудова відсікання при розв’язанні цілочислової лінійної задачі за методом віток і границь?
9. Опишіть схему розв’язання цілочислової задачі лінійного програмування за методом віток і границь.
10. За рахунок яких перетворень вдається побудувати сполучений базис при додаванні відсікаючого обмеження?

ТЕМА 8
ЗАДАЧІ НЕЛІНІЙНОГО ПРОГРАМУВАННЯ. ОСНОВНІ МЕТОДИ ЇХ РОЗВ’ЯЗАННЯ І АНАЛІЗУ

8.1 Постановка задачі нелінійного програмування (ЗНП)

Залежності між керованими змінними далеко не завжди можна описати за допомогою адекватної лінійної моделі. Наприклад, у лінійних моделях ціну товару вважають незалежною від кількості реалізованого продукту, у той час як вона може залежати від обсягу партії товару. Щодо технологічних обмежень можна помітити, що витрата певних видів сировини і ресурсів відбувається не лінійно, а стрибкоподібно (залежно від обсягу виробництва). Спроби врахувати ці фактори призводять до формулювання загальніших і складніших оптимізаційних задач. Вивчення методів їх розв’язання складає предмет наукової області, що одержала назву нелінійного програмування.

Для задач нелінійного програмування не існує універсального методу розв’язання, тому щораз необхідно доводити існування розв’язку задачі, а також його одиничність. Відомі точні методи розв’язання нелінійних задач, але їх алгоритми є трудомісткими навіть для сучасного програмного забезпечення ЕОМ. На практиці частіше користуються наближеними методами, проблема яких пов'язана з пошуком локальних і глобальних оптимумів. Більшість наближених методів дозволяють визначити локальний оптимум. Визначивши всі локальні оптимуми та порівнявши їх, можна знайти глобальний оптимум. Але такий підхід є неефективним для практичних розрахунків. Слід зазначити, що якщо в задачах лінійного програмування оптимальний розв’язок завжди знаходився на границі області обмежень, то у задачі нелінійного програмування він може знаходитися також і усередині цієї області.

У класичній теорії оптимізації для пошуку точок максимуму та мінімуму (екстремальних точок) функцій як за відсутності так і за наявності обмежень на змінні, використовують апарат диференціального обчислення. Екстремальна точка функції $f(x)$ визначає або її максимальне, або мінімальне значення. З математичної точки зору точка $x_0=(x_1, x_2, \ldots, x_n)$ є точкою максимуму, якщо значення функції f в оточенні точки x_0 не перевищують $f(x_0)$. На рисунку 8.1 показані точки максимуму та мінімуму функції однієї змінної $f(x)$. Точки x_1, x_2, x_3, x_4 і x_6 складають множину екстремальних точок функції $f(x)$. Причому точка x_6 є точкою **глобального** (абсолютного) максимуму, тому що $f(x) = \max\{f(x_1), f(x_3), f(x_6)\}$, а точки $f(x_1)$ і $f(x_3)$ – точками **локального** (відносного) максимуму. Необхідною умовою існування екстремуму є рівність нулю похідних від $f(x)$. Але похідні дорівнюють нулю також і в точках перегину функції $f(x)$ у випадку однієї змінної (точка x_5), і в сідлових точках у випадку функції двох змінних. Тому рівність нулю похідних від $f(x)$ є необхідною, але недостатньою умовою існування екстремуму, а точки, у яких виконується ця умова, називають **стаціонарними**.

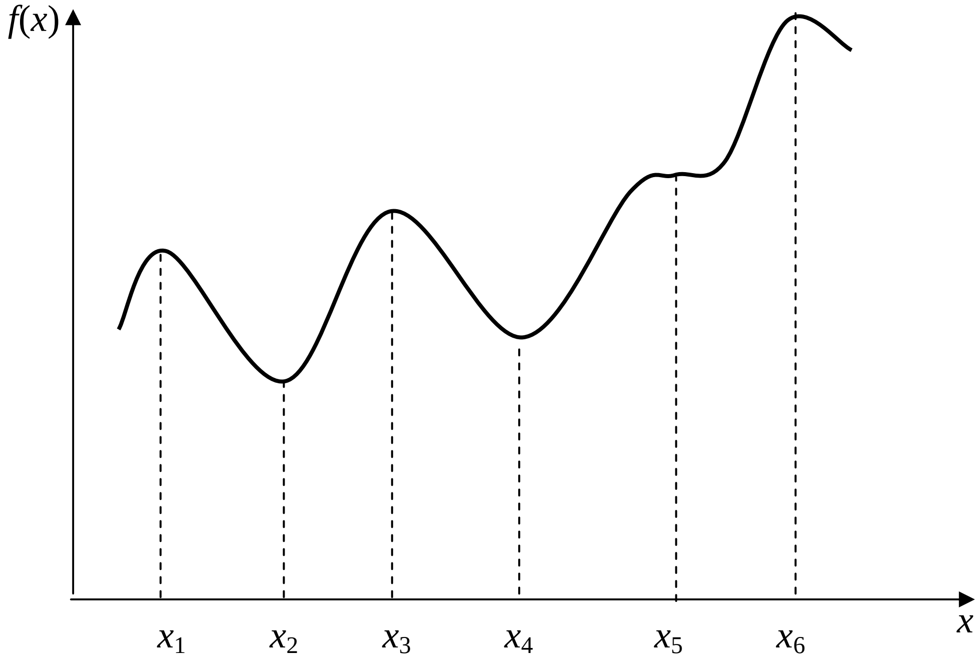


Рисунок 8.1 – Точки максимуму і мінімуму функції

Більшість методів, використовуваних для розв'язання задач нелінійного програмування, базуються на теорії диференціального обчислення. Серед них розрізняють **прямі** та **непрямі** методи. За прямими методами пошук оптимуму ведеться в напрямку найскорішого зростання або убування цільової функції. До таких методів належать градієнтні методи, зокрема, метод найшвидшого спуску і метод припустимих напрямків. Непрямі методи припускають перетворення вихідної задачі до виду, що дозволяє спростити пошук екстремуму цільової функції. До них належить квадратичне програмування та сепарабельне програмування, а також геометричне та стохастичне програмування.

Загальну задачу нелінійного програмування (ЗЗНП) визначають як задачу із знаходження максимуму (або мінімуму) цільової функції $f(x_1, x_2, ..., x_n)$ на множині D, зумовленій системою обмежень

$$\begin{gathered} g_1(x_1, x_2, ..., x_n) \le 0, \\ g_2(x_1, x_2, ..., x_n) \le 0, \\ \dots\dots\dots\dots\dots \\ g_i(x_1, x_2, ..., x_n) \le 0, \\ g_{i+1}(x_1, x_2, ..., x_n) \le 0, \\ \dots\dots\dots\dots\dots \\ g_m(x_1, x_2, ..., x_n) = 0, \\ x_j \ge 0, \quad j = \overline{1, n}, \end{gathered} \tag{8.1}$$

де хоча б одна з функцій f або g_i є нелінійною.

Очевидно, що питання про тип оптимізації не є принциповим. Тому, для визначеності, надалі розглядатимемо задачі максимізації.

Як і в ЗЛП, вектор $x^*=(x_1, x_2, ..., x_n)\in D$ називають припустимим планом, а якщо для будь-якого $x\in D$ виконується нерівність $f(x^*)>f(x)$, то x^* називають оптимальним планом. У цьому випадку x^* є точкою глобального максимуму.

З погляду економічної інтерпретації $f(x)$ може розглядатися як дохід, що одержує підприємство при плані випуску x, а $g_i(x)<0$ як технологічні обмеження на можливості випуску продукції.

Набір обмежень, які визначають множину D, при необхідності завжди можна звести або до системи, що складається з одних нерівностей, або, додавши фіктивні змінні, до системи рівнянь. Перелічимо властивості ЗНП, які істотно ускладнюють процес їх розв’язання порівняно із задачами лінійного програмування.

1. Множина припустимих планів D може мати дуже складну структуру (наприклад, бути неопуклою або незв'язною).

2. Глобальний максимум (мінімум) може досягатися як усередині множини D, так і на її границях (де він, загалом кажучи, не збігатиметься з жодним з локальних екстремумів).

3. Цільова функція f може бути недиференційованою, що утруднює застосування класичних методів математичного аналізу.

У чинність названих факторів задачі нелінійного програмування настільки різноманітні, що для них не існує загального методу розв’язання.

8.2 Класичний метод оптимізації з використанням множників Лагранжа

Одним з найзагальніших підходів до розв’язання задач з пошуку екстремуму функції за наявності сполучних обмежень на її змінні (задача умовної оптимізації) є **метод Лагранжа**. Багатьом студентам він відомий з курсу диференціального обчислення. Ідея даного методу полягає у зведенні задачі пошуку умовного екстремуму цільової функції

$$f(x_1, x_2, ..., x_n) \tag{8.2}$$

на множині припустимих значень D, описуваній системою рівнянь

$$\begin{gathered} g_1(x_1, x_2, ..., x_n)= 0, \\ g_2(x_1, x_2,..., x_n)= 0, \\ \dots\dots\dots\dots \\ g_m(x_1, x_2, ..., x_n)= 0, \end{gathered} \tag{8.3}$$

до задачі безумовної оптимізації функції

$$\Phi(x, \lambda) = f(x_1, x_2, ..., x_n)+ \lambda_1 g_1(x_1, x_2, ..., x_n) + ...+ \lambda_m g_m(x_1, x_2, ..., x_n), \tag{8.4}$$

де λ_i - вектор додаткових змінних, називаних невизначеними множниками Лагранжа.

Рівняння (8.3) називають рівняннями зв'язку, а функцію $\Phi(x, \lambda)$ - функцією Лагранжа.

Для функції $\Phi(x, \lambda)$ вирішують систему рівнянь

$$\begin{cases} \dfrac{\partial \Phi(x,\lambda)}{\partial x_j} = 0, (j = \overline{1,n}), \\ \dfrac{\partial \Phi(x,\lambda)}{\partial \lambda_i} = g_i(x) = 0, (i = \overline{1,m}) \end{cases} \tag{8.5}$$

щодо змінних x і λ.

Метод Лагранжа складається з наступних етапів.

1. Складання функції Лагранжа $\Phi(x,\lambda)$.
2. Знаходження частинних похідних

$$\frac{\partial \Phi(x,\lambda)}{\partial x_j},\ j = \overline{1,n} \quad \text{і} \quad \frac{\partial \Phi(x,\lambda)}{\partial \lambda_i},\ i = \overline{1,m}.$$

3. Розв’язання системи рівнянь (8.5) щодо змінних x і λ.
4. Дослідження точок, що задовольняють системі (8.5), на максимум (мінімум) за допомогою достатньої ознаки екстремуму.

Наявність останнього (четвертого) етапу пояснюється тим, що розглянутий алгоритм виконує необхідну, але не достатню умову екстремуму. Для визначення достатніх ознак умовного екстремуму і його типу існує спеціальний алгоритм, як правило, важко застосовний на практиці.

Основне практичне значення методу Лагранжа полягає в тому, що він дозволяє перейти від умовної оптимізації до безумовної. Проте задача розв’язання системи рівнянь (8.5), до якої збігається даний метод, у загальному випадку не простіше вихідної проблеми пошуку екстремуму (8.2)-(8.3). Методи, що припускають таке розв’язання, називають непрямими. Їх можна застосовувати для досить вузького класу задач, для яких вдається одержати лінійну або таку, що збігається до лінійної систему рівнянь (8.5). Їх застосування пояснюється необхідністю одержати розв’язок екстремальної задачі в аналітичній формі. При розв’язанні конкретних практичних задач зазвичай використовують прямі методи, засновані на ітеративних процесах обчислення та порівнянні значень функцій, що оптимізують.

8.3 Опукле програмування

Основний недолік методів нелінійного програмування полягає в тому, що за їхньою. допомогою не вдається знайти глобальний екстремум за наявності кількох локальних екстремумів. Теоретично нелінійне програмування розроблене тільки для одного окремого випадку опуклих функцій, і відповідно цей розділ названий **опуклим програмуванням**.

Функцію $f(x_1, x_2, ..., x_n)$ називають опуклою в області D, якщо для будь-яких двох точок $x^{(1)}, x^{(2)} \in D$ *і будь-якого* $\lambda = \overline{1,0}$ *виконується нерівність*

$$f((1-\lambda)x^{(1)} + \lambda x^{(2)}) \le (1-\lambda) f(x^{(1)}) + \lambda f(x^{(2)}), \tag{8.6}$$

якщо

$$f((1-\lambda)x^{(1)} + \lambda x^{(2)}) \ge (1-\lambda) f(x^{(1)}) + \lambda f(x^{(2)}), \tag{8.7}$$

то функцію називають увігнутою.

Геометричний зміст понять опуклості та увігнутості для функції однієї змінної представлений на рисунку 8.2. З нього видно, що графік опуклої функції лежить нижче відрізка, що з'єднує точки ($x^{(1)}$, $f(x^{(1)})$) і ($x^{(2)}$, $f(x^{(2)})$), а графік увігнутої - вище.

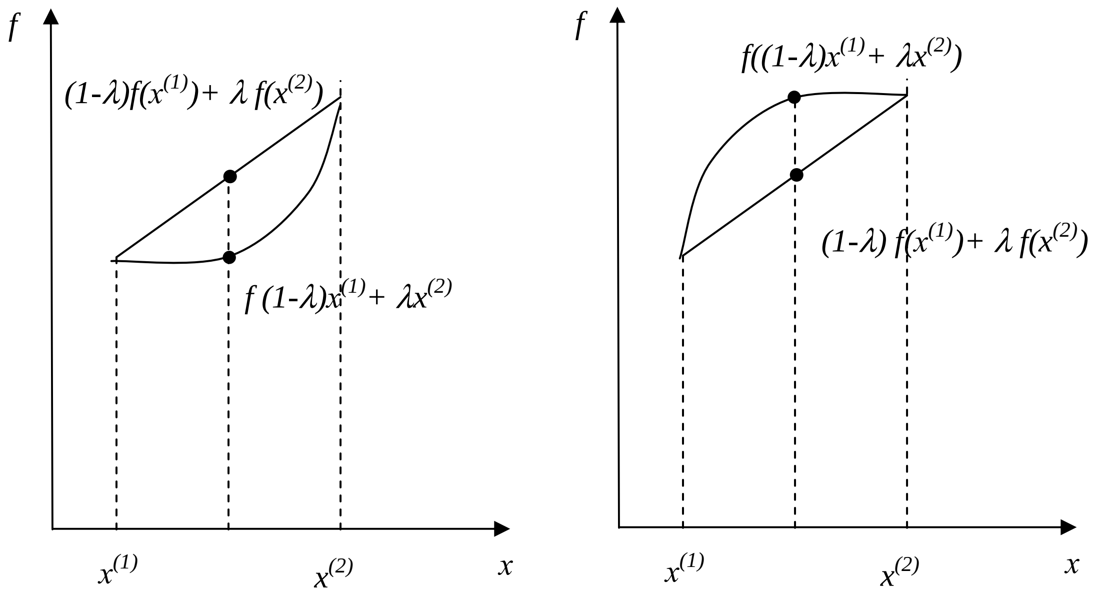


Рисунок 8.2 - Графіки опуклої та увігнутої функцій

Можна довести, що достатньою умовою опуклості функції $f(x_1, x_2, ..., x_n)$ є позитивна визначеність матриці

$$H(x) = \left(\frac{\partial^2 f(x)}{\partial x_i \partial x_j} \right)_{n \times n}, \quad (8.8)$$

називаної матрицею Гессе, у всіх точках $x \in D$. Відповідно, достатньою умовою увігнутості є негативна визначеність матриці Гессе. Зокрема, для функцій однієї змінної достатньою умовою опуклості (увігнутості) є виконання нерівності $f''(x) \geq 0$ ($f''(x) \leq 0$).

Як видно з геометричної інтерпретації, для опуклої функції локальний екстремум, якщо він існує, збігається з глобальним. Для функції багатьох змінних точка x_0 являє собою вектор $f'(x)$ − вектор перших похідних (градієнт) функції $f(x)$, а матриця Гессе (гессіан) $f''(x)$ − симетричну матрицю других частинних похідних функції $f(x)$. Характер стаціонарної точки x_0 пов'язаний зі знаковизначеністю матриці Гессе $f''(x^*)$.

Знаковизначеність матриці A залежить від знаків квадратичної форми

$$Q(x) = X^T A X = \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} x_i x_j \ . \quad (8.9)$$

Квадратична форма $Q(x)$ є позитивно визначеною (напіввизначеною), якщо значення всіх кутових мінорів визначника $|A|$ додатні (невід'ємні). У цьому випадку матрицю A називають позитивно визначеною (напіввизначеною). k-м кутовим мінором визначника матриці $A_{n\times n}$ називають визначник виду

$$\begin{vmatrix} a_{11}\ a_{12} \dots a_{1k} \\ a_{21}\ a_{22} \dots a_{2k} \\ \vdots \quad \vdots \quad \dots \vdots \\ a_{k1}\ a_{k2} \dots a_{kk} \end{vmatrix}, \quad k = \overline{1,n}. \tag{8.10}$$

Квадратична форма $Q(x)$ є негативно визначеною, якщо значення k-х кутових мінорів визначника $|A|$ відмінні від нуля і мають знак $(-1)^k$. У цьому випадку матрицю A називають негативно визначеною.

Квадратична форма $Q(x)$ є негативно напіввизначеною, якщо значення k-х кутових мінорів визначника $|A|$ дорівнюють нулю або мають знак $(-1)^k$.

Розглянемо приклад. Нехай є функція

$$f(x_1, x_2, x_3) = x_1 + 2x_3 + x_2 x_3 - x_1^2 - x_2^2 - x_3^2,$$

необхідно визначити її екстремум. Визначимо градієнт функції $f(x)$ (необхідна умова екстремуму)

$$\frac{\partial f}{\partial x_1} = 1 - 2x_1 = 0,$$

$$\frac{\partial f}{\partial x_2} = x_1 - 2x_2 = 0,$$

$$\frac{\partial f}{\partial x_3} = 2 + x_2 - 2x_3 = 0,$$

звідки одержимо стаціонарну точку $x_0 = \left(\frac{1}{2}, \frac{2}{3}, \frac{4}{3}\right)$. Визначимо характер стаціонарної точки. Для цього перевіримо достатність умови, склавши матрицю Гессе

$$H(x_0) = \begin{pmatrix} \frac{\partial^2 f}{\partial^2 x_1} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \frac{\partial^2 f}{\partial x_1 \partial x_3} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial^2 x_2} & \frac{\partial^2 f}{\partial x_2 \partial x_3} \\ \frac{\partial^2 f}{\partial x_3 \partial x_1} & \frac{\partial^2 f}{\partial x_3 \partial x_2} & \frac{\partial^2 f}{\partial^2 x_3} \end{pmatrix}_{x_0} = \begin{pmatrix} -2 & 0 & 0 \\ 0 & -2 & 1 \\ 0 & 1 & -2 \end{pmatrix}$$

і кутові мінори матриці $H(x_0)$

M_1= -2; M_2=(-2)*(-2)-0*0=4; M_3=-2*[(-2)*(-2)-1*1]=-6.

Отже, матриця $H(x_0)$ є негативно визначеною, звідки випливає, що функція $f(x)$ – увігнута, а стаціонарна точка $x_0 = \left(\frac{1}{2}, \frac{2}{3}, \frac{4}{3}\right)$ є точкою максимуму.

8.4 Необхідні й достатні умови існування сідлової точки. Теорема Куна-Таккера

Зупинимося на певних фундаментальних моментах теорії нелінійного програмування. Відправною точкою для них є поширення методу Лагранжа на розв'язання ЗНП із обмеженнями у формі нерівностей:

$$\begin{gathered} f(x_1, x_2, ..., x_n) \to \max, \\ g_1(x_1, x_2, ..., x_n) \le 0, \\ g_2(x_1, x_2, ..., x_n) \le 0, \\ \ldots\ldots\ldots\ldots\ldots\ldots \\ g_m(x_1, x_2, ..., x_n) \le 0. \end{gathered} \tag{8.11}$$

Визначимо функцію Лагранжа:

$$\Phi(x,\lambda) = f(x_1, x_2, \ldots, x_n) - \sum_{i=1}^{m} \lambda_i g_i(x_1, x_2, \ldots, x_n). \tag{8.12}$$

Пару векторів (x, λ) називають сідловою точкою функції Φ(x, λ) у певній області, якщо для будь-яких x і λ, що належать цієї області,

$$\Phi(x,\lambda^*) \le \Phi(x^*,\lambda^*) \le \Phi(x^*,\lambda). \tag{8.13}$$

Нерівності (8.13) називають нерівностями сідлової точки.

Як приклад розглянемо функцію $\Phi(x,\lambda) = -x^2 + \lambda^2$. Сідловою точкою для цієї функції є точка (0, 0). Справді, $\Phi(0,0) = 0$, $\Phi(x,0) =-x^2$, $\Phi(0, \lambda)=\lambda^2$, а для будь-яких x і λ виконуються нерівності $-x^2 \le 0$ і $0 \le \lambda^2$.

На рисунку 8.3 зображений графік функції $\Phi(x,\lambda)$ (гіперболічний параболоїд), і, як видно, в оточенні точки (0,0) він дійсно за формою нагадує сідло, чим і пояснюється походження відповідного терміну.

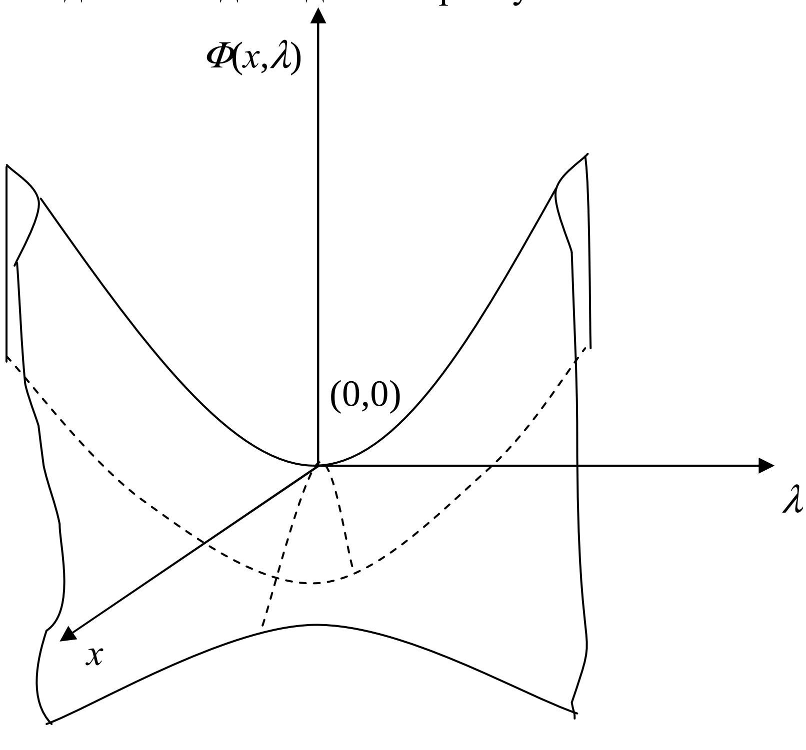


Рисунок 8.3 - Графік функції $\Phi(x,\lambda)$

Центральне місце в теорії нелінійного програмування займає **теорема Куна-Таккера**, що пов'язує розв’язання ЗНП із наявністю сідлової точки у відповідної функції Лагранжа.

Теорема 8.1. (Достатня умова екстремуму).

Якщо (x^*, λ^*) *- сідлова точка функції Лагранжа в області* $x \in X \supseteq D$, $\lambda \geq 0$, *то* x^* *є оптимальним планом задачі* (8.11), причому справедливе так зване правило додаткової нетвердості (умова Слейтера):

$$\sum_{i=1}^{m} \lambda_i^* g_i(x^*) = 0. \qquad (8.14)$$

Доказ.

Скористаємося визначенням сідлової точки та запишемо

$$f(x) - \sum_{i=1}^{m} \lambda_i^* g_i(x) \leq f(x^*) - \sum_{i=1}^{m} \lambda_i^* g_i(x^*) \leq f(x^*) - \sum_{i=1}^{m} \lambda_i g_i(x^*) \qquad (8.15)$$

при всіх $x \in X$, $\lambda \geq 0$. З другої нерівності в (8.15) випливає, що

$$\sum_{i=1}^{m} \lambda_i g_i(x^*) \leq \sum_{i=1}^{m} \lambda_i^* g_i(x^*) \text{ для } \lambda \geq 0. \qquad (8.16)$$

Проте (8.16) може мати місце тільки тоді, коли $g(x^*) \leq 0$ при всіх $i = \overline{1,m}$. Дійсно, якщо існує таке k, що $g_k(x^*) > 0$, то поклавши $\lambda_i = 0$ для всіх $i \neq k$ і вибравши досить велике $\lambda_k > 0$, можна домогтися того, що значення $\sum_{i=1}^{m} \lambda_i g_i(x^*) = \lambda_k g_k(x^*)$ опиниться більшим за постійний вираз $\sum_{i=1}^{m} \lambda_i^* g_i(x^*)$.

З того, що для всіх $i = \overline{1,m}$ виконуються нерівності $g_i(x^*) \leq 0$, випливає, що x^* є припустимим планом задачі (8.11).

Якщо до лівої частини нерівності (8.16) підставити значення $\lambda_i = 0$ для всіх $i = \overline{1,m}$, одержимо

$$\sum_{i=1}^{m} \lambda_i^* g_i(x^*) \geq 0.$$

Разом з тим з того що, $g_i(x^*) \leq 0$ і $\lambda_i^* \geq 0$ випливає оцінка

$$\sum_{i=1}^{m} \lambda_i^* g_i(x^*) \leq 0.$$

Спільний розгляд останніх двох нерівностей призводить до правила додаткової нетвердості у точці x^*:

$$\sum_{i=1}^{m} \lambda_i^* g_i(x^*) = 0.$$

Тоді на підставі лівої частини нерівності сідлової точки (8.15) маємо, що для всіх $x \in X$ (у тому числі й для $x \in D$)

$$f(x^*) \geq f(x) - \sum_{i=1}^{m} \lambda_i^* g_i(x).$$

За умовою ЗНП для будь-яких $x \in D$ вірні нерівності $g_i(x^*) \leq 0$, що у сполученні з умовою $\lambda_i^* = 0$ дозволяє записати

$$\sum_{i=1}^{m} \lambda_i^* g_i(x) \leq 0.$$

Виходить,

$$f(x) - \sum_{i=1}^{m} \lambda_i^* g_i(x) \geq f(x).$$

Остаточно одержуємо, що для будь-яких $x \in D$ справедливе співвідношення $f(x^*) \geq f(x)$, тобто x^* - оптимальний план задачі (8.11).

Значення теореми Куна-Таккера полягає в тому, що вона дозволяє зв'язати процес розв'язання оптимізаційної задачі з пошуком сідлових точок функції Лагранжа, тобто з максимізацією цієї функції за x і мінімізацією за λ.

Нехай $F(x)$ є функцією, що ставить у відповідність кожному значенню x мінімальне значення функції $\Phi(x,\lambda)$ за λ:

$$F(x) = \min_{\lambda \geq 0} \Phi(x, \lambda)$$

і за аналогією

$$G(\lambda) = \max_{x \in X} \Phi(x, \lambda).$$

Розглянемо задачу відшукання максимуму функції $F(x)$

$$F(x) = \min_{\lambda \geq 0} \Phi(x, \lambda) \to \max, \quad x \in X \qquad (8.17)$$

і задачу мінімізації $G(\lambda)$

$$G(\lambda) = \max_{x \in X} \Phi(x, \lambda) \to \min, \quad \lambda \geq 0. \qquad (8.18)$$

Очевидно, що

$$F(x) = \min_{\lambda \geq 0} \Phi(x, \lambda) = \begin{cases} f(x), \; x \in D, \\ \infty, \quad x \notin D \end{cases}.$$

Звідси випливає, що максимум $F(x)$ знаходиться в припустимій області D і збігається з максимумом цільової функції $f(x)$ задачі (8.11):

$$\max_{x \in X} F(x) = \max_{x \in X} \min_{\lambda \geq 0} \Phi(x, \lambda) = \max_{x \in D} f(x).$$

Отже, задача (8.17), у певному змісті, рівносильна (8.11). Аналогічні висновки можна отримати і для (8.18). Задачі (8.17) і (8.18) утворюють двоїсту пару. Це відношення є узагальненням відносини подвійності для задач лінійного програмування. Відповідно, за певних умов пара двоїстих задач нелінійного програмування має властивості, аналогічні властивостям двоїстих лінійних задач. Зокрема, при будь-яких $x \in X$, $\lambda \geq 0$

$$F(x) \leq G(\lambda). \qquad (8.19)$$

Умова (8.19) знаходить широке застосування при побудові оцінок в ітеративних методах розв'язання оптимізаційних задач. Наприклад, якщо є можливість приблизно вирішити пряму і двоїсту задачі і одержати послідовності наближень, то за допомогою нерівностей виду

$$f(x^{(k)}) \leq f(x^*) \leq G(\lambda^{(k)})$$

можна визначити момент зупинки обчислювальної процедури.

Можливим є варіант виведення виразів для цільових функцій і обмежень пари двоїстих задач лінійного програмування із загального визначення відносини подвійності для нелінійних задач. Також відзначимо, що в процесі формування нелінійних двоїстих задач існує велика неоднозначність: їх вид можна варіювати, включаючи до множини X частину обмежень $g_i(x) \leq 0$.

8.5 Деякі методи розв'язання задач НЛП

8.5.1 Градієнтні методи розв'язання задач безумовної оптимізації

Провідне місце серед прямих методів розв'язання екстремальних задач займає градієнтний метод (точніше, сімейство градієнтних методів) пошуку стаціонарних точок диференційованої функції. Нагадаємо, що стаціонарною називають точку, у якій $\nabla f(x)=0$ і яка відповідно до необхідної умови оптимальності є «підозрілою» на наявність локального екстремуму. Застосовуючи градієнтний метод, знаходять множину точок локальних максимумів (або мінімумів), серед яких визначають максимум (або мінімум) глобальний.

Ідея цього методу заснована на тому, що градієнт функції вказує напрямок її найскорішого зростання в оточенні тієї точки, в якій він обчислений. Тому, якщо з певної поточної точки $x^{(1)}$ переміщатися в напрямку вектора $\nabla f(x^{(1)})$, функція f зростатиме, принаймні, у певному оточенні $x^{(1)}$. Отже, для точки $x^{(2)} = x^{(1)} + k\nabla f(x^{(1)})$, ($k>0$), що лежить у такому оточенні, справедлива нерівність $f(x^{(1)}) \leq f(x^{(2)})$. Продовжуючи цей процес, ми поступово наближатимемось до точки певного локального максимуму.

Однак як тільки визначається напрямок руху, відразу встає питання, як далеко треба рухатися в цьому напрямку, тобто виникає проблема вибору кроку r у рекурентній формулі

$$x^{(k+1)} = x^{(k)} + r\nabla f(x^{(k)}), \qquad (8.20)$$

що задає послідовність точок, які прагнуть до точки максимуму.

Залежно від способу її розв'язання розрізняють різні варіанти градієнтного методу. Зупинимося на найвідоміших з них.

Метод найскорішого спуску. Назву методу можна було б розуміти буквально, якби мова йшла про мінімізацію цільової функції. Проте, за традицією таку назву використовують і при розв'язанні задачі на максимум.

Нехай $f(x) = f(x_1, x_2, ..., x_n)$ - диференційована функція, а $x^{(k)}=(x_1^{(k)}, x_2^{(k)}, ..., x_n^{(k)})$ - певна поточна точка. Будь-яких загальних рекомендацій щодо вибору вихідної точки (початкового наближення) $x^{(0)}$ не існує, але за можливістю вона має знаходитися близько від шуканого оптимального плану x^*. Якщо $x^{(k)}$ - нестаціонарна точка, то при русі в напрямку $\nabla f(x^{(k)})$ функція $f(x)$ на певному проміжку обов'язково зростатиме. Звідси виникає необхідність такого вибору кроку, щоб рух у зазначеному напрямку тривало доти, поки зростання не припиниться. Виразимо залежність значення $f(x)$ від крокового множника $r > 0$, покладаючи $x = x^{(k)} + r\nabla f(x^{(k)})$

$$f(x) = f(x^{(k)} + r\nabla f(x^{(k)})) = \varphi(r), \tag{8.21}$$

або в координатній формі,

$$\varphi(r) = f\left(x_1^{(k)} + r\frac{\partial f(x^{(k)})}{\partial x_1}, \ldots, x_n^{(k)} + r\frac{\partial f(x^{(k)})}{\partial x_n}\right). \tag{8.22}$$

Щоб домогтися найбільшого з можливих значень f при русі у напрямку $\nabla f(x^{(k)})$, потрібно вибрати таке значення r, що максимізує функцію $\varphi(r)$. Для обчислення r використовують необхідну умову екстремуму $\frac{d\varphi(r)}{dr} = 0$. Помітимо, що якщо для будь-якого $r>0$ $\frac{d\varphi(r)}{dr} > 0$, то функція $f(x)$ не обмежена зверху (тобто не має максимуму). У протилежному випадку, на підставі (8.22) одержуємо

$$\frac{d\varphi(r)}{dr} = \frac{\partial f(x)}{\partial x_1}\times\frac{dx_1}{dr} + \cdots + \frac{\partial f(x)}{\partial x_n}\times\frac{dx_n}{dr}, \tag{8.23}$$

що, у свою чергу, дає

$$\frac{d\varphi(r)}{dr} = \frac{\partial f(x)}{\partial x_1}\times\frac{\partial f(x^{(k)})}{\partial x_1} + \cdots + \frac{\partial f(x)}{\partial x_n}\times\frac{\partial f(x^{(k)})}{\partial x_n} = \nabla f(x)\nabla f(x^{(k)}). \tag{8.24}$$

Якщо вважати, що наступна точка $x^{(k+1)}$ відповідає оптимальному значенню, то в ній повинна виконуватися умова $\frac{d\varphi(r)}{dr} = 0$, і r треба знаходити з умови $\nabla f(x^{(k+1)})\nabla f(x^{(k)}) = 0$ або

$$\nabla f(x^{(k)} + r\nabla f(x^{(k)}))\nabla f(x^{(k)}) = 0. \tag{8.25}$$

Умова (8.25) означає рівність нулю скалярного добутку градієнтів функції f в точках $x^{(k+1)}$ і $x^{(k)}$. Геометрично її можна інтерпретувати як перпендикулярність векторів градієнтів функції f у зазначених точках. Продовжуючи геометричну інтерпретацію методу найскорішого спуска, відзначимо, що в точці $x^{(k+1)}$ вектор $\nabla f(x^{(k+1)})$, будучи градієнтом, перпендикулярний до лінії рівня, що проходить через цю точку. Відповідно вектор $\nabla f(x^{(k)})$ є дотичним до цієї лінії. Отже, рух у напрямку градієнта $\nabla f(x^{(k)})$ треба продовжувати доти, поки він перетинає лінії рівня функції, що оптимізують.

Після того як точку $x^{(k+1)}$ знайдено, вона стає поточною для чергової ітерації. На практиці ознакою досягнення стаціонарної точки служить досить мала зміна координат точок, розглянутих на послідовних ітераціях. Одночасно із цим координати вектора $\nabla f(x^{(k)})$ повинні наближатись до нуля.

Метод дроблення кроку. Для знаходження кроку r у методі найскорішого спуска потрібно вирішити рівняння (8.25), що може виявитися досить складним. Тому часто обмежуються «підбором» такого значення r, що $\varphi(r) > \varphi(0)$. Для цього задаються певним початковим значенням r_1 (наприклад, $r_1=1$) і перевіряють умову $\varphi(r_1) > \varphi(0)$. Якщо вона не виконується, то покладають

$$r_2 = \frac{r_1}{2}$$

і т. д. доти, поки не вдається знайти підходящий крок, з яким переходять до наступної точки $x^{(k+1)}$. Критерій завершення алгоритму такий самий, як і в методі найскорішого спуску.

8.5.2 Квадратичне програмування (КП)

До задач квадратичного програмування належить спеціальний клас задач нелінійного програмування, для яких цільова функція $f(x)$ квадратична, а всі обмеження - лінійні. Застосувавши до цієї задачі теорему Куна-Таккера, одержують умову для оптимального розв'язку у вигляді системи лінійних рівнянь, вирішити які можна за симплекс-методом. У матричному виді задачу квадратичного програмування формулюють в такий спосіб:

Максимізувати (мінімізувати) цільову функцію $f = \overline{C}\,\overline{X} + \overline{X}^T\,\overline{D}\,\overline{X}$ при обмеженнях

$$\overline{A}\,\overline{X} \le b, \quad \overline{X} \ge 0, \tag{8.26}$$

де

$$\overline{X} = (x_1, x_2, ..., x_n)^T,$$

$$\overline{C} = (c_1, c_2, ..., c_n),$$

$$b=(b_1, b_2, \ldots, b_m)^T,$$

$$\overline{A} = \begin{pmatrix} a_{11}\, a_{12} \ldots a_{1n} \\ a_{21}\, a_{22} \ldots a_{2n} \\ \vdots \quad \vdots \quad \ldots \vdots \\ a_{m1}\, a_{m2} \ldots a_{mn} \end{pmatrix},$$

$$\overline{D} = \begin{pmatrix} d_{11}\, d_{12} \ldots d_{1n} \\ d_{21}\, d_{22} \ldots d_{2n} \\ \vdots \quad \vdots \quad \ldots \vdots \\ d_{n1}\, d_{n2} \ldots d_{nn} \end{pmatrix}.$$

Функція $\overline{X}^T\,\overline{D}\,\overline{X}$, де D – симетрична матриця, що є квадратичною формою (8.9). Матриця D є негативно визначеною в задачі максимізації й позитивно визначеною – у задачі мінімізації. Це означає, що функція f є строго опуклою за змінними $\overline{X}$ у задачі мінімізації й строго увігнутою – у задачі максимізації.

Обмеження задачі є лінійними, що гарантує опуклість області припустимих розв’язків.

Складемо функцію Лагранжа для задачі (8.26)

$$\Phi(\overline{X},\overline{\lambda}) = \overline{C}\,\overline{X} + \overline{X}^T\,\overline{D}\,\overline{X} + \overline{\lambda}(b - \overline{A}\,\overline{X}). \quad (8.27)$$

Відповідно до теореми Куна-Таккера для існування сідлової точки необхідно й достатньо, щоб похідні функції Лагранжа за x_j були меншими за нуль, а за λ_i – перевищували б нуль, маємо:

$$\frac{\partial \Phi(\overline{X},\overline{\lambda})}{\partial x_j} = \overline{C} + 2\overline{D}\,\overline{X} - \overline{\lambda}\,\overline{A} \le 0, \quad (8.28)$$

причому, якщо $\dfrac{\partial \Phi}{\partial x_j} < 0$, відповідні x_j дорівнюють нулю;

$$\frac{\partial \Phi(\overline{X},\overline{\lambda})}{\partial \lambda_i} = b - \overline{A}\,\overline{X} \ge 0, \quad (8.29)$$

якщо $\dfrac{\partial \Phi}{\partial \lambda_i} > 0$, відповідні λ_i дорівнюють нулю.

Щоб нерівності (8.28) і (8.29) привести до рівностей, введемо два допоміжних вектори: $\overline{V} = (v_1, v_2, \ldots, v_n) \ge 0$ і $\overline{S} = (s_1, s_2, \ldots, s_m) \ge 0$, причому виберемо $v_j>0$, якщо $\dfrac{\partial \Phi}{\partial x_j} < 0$ й $v_j=0$, якщо $\dfrac{\partial \Phi}{\partial x_j} = 0$, а також $\lambda_i>0$, якщо $\dfrac{\partial \Phi}{\partial \lambda_i} > 0$ й $\lambda_i=0$, якщо $\dfrac{\partial \Phi}{\partial \lambda_i} = 0$, одержимо

$$\overline{C} + 2\overline{D}\,\overline{X} - \overline{\lambda}\,\overline{A} + \overline{V} = 0 \quad (8.30)$$

$$b - \overline{A}\,\overline{X} - \overline{S} = 0. \quad (8.31)$$

Порівнявши компоненти векторів $\overline{X}$ і $\overline{V}$, а також $\overline{\lambda}$ і $\overline{S}$, одержимо дві умови додаткової нетвердості

$$\overline{X}^T\,\overline{V} = 0 \text{ і } \overline{\lambda}^T\,\overline{S} = 0. \quad (8.32)$$

З умов (8.32) випливає, що хоча б n змінних з $\overline{X}$ і $\overline{V}$, а також m змінних з $\overline{\lambda}$ і $\overline{S}$ обертаються на нуль. Якщо існує оптимальний розв’язок задачі (8.26), то він є одним з базисних розв’язків системи (8.30), (8.31). Для знаходження припустимого базисного розв’язку використовується симплекс-метод.

Розглянемо приклад. Знайти

$$\max f(x_1, x_2) = \max(10x_1 + 20x_2 + x_1x_2 - 2x_1^2 - 2x_2^2)$$

при обмеженнях:

$$8 - x_2 \ge 0,$$
$$10 - x_1 - x_2 \ge 0,$$
$$x_1 \ge 0;\ x_2 \ge 0.$$

Можна показати, що функція $f(x_1, x_2)$ є увігнутою функцією. Складемо функцію Лагранжа

$$\Phi(x_1, x_2, \lambda_1, \lambda_2) = 10x_1 + 20x_2 + x_1x_2 - 2x_1^2 - 2x_2^2 + \lambda_1(8 - x_2) + \lambda_2(10 - x_1 - x_2).$$

Застосувавши теорему Куна-Таккера, одержимо умови існування сідлової точки:

$$\frac{\partial \Phi}{\partial x_1} = 10 + x_2 - 4x_1 + \lambda_2 \le 0,$$

$$\frac{\partial \Phi}{\partial x_2} = 20 + x_1 - 4x_2 - \lambda_1 - \lambda_2 \le 0,$$

$$\frac{\partial \Phi}{\partial \lambda_1} = 8 - x_2 \ge 0,$$

$$\frac{\partial \Phi}{\partial \lambda_2} = 10 - x_1 - x_2 \ge 0.$$

Одержали систему лінійних нерівностей:

$$\begin{cases} 10 + x_2 - 4x_1 + \lambda_2 \le 0, \\ 20 + x_1 - 4x_2 - \lambda_1 - \lambda_2 \le 0, \\ 8 - x_2 \ge 0, \\ 10 - x_1 - x_2 \ge 0. \end{cases}$$

Введемо вільні змінні $\overline{V}$ та $\overline{S}$, що обертають систему нерівностей на систему лінійних рівнянь, вирішивши яку за симплекс-методом, знайдемо оптимум

$$\begin{cases} 4x_1 - x_2 - \lambda_2 - v_1 = 10, \\ -x_1 + 4x_2 + \lambda_1 + \lambda_2 - v_2 = 20, \\ x_2 + s_1 = 8, \\ x_1 + x_2 + s_2 = 10. \end{cases}$$

Для знаходження припустимого базисного плану отриманої задачі скористаємося методом мінімізації нев'язань. Для цього побудуємо допоміжну задачу, увівши до обмежень 1 і 2 фіктивні змінні x_3 і x_4, які входять до цільової функції допоміжної задачі з коефіцієнтами M (а інші змінні – з нульовими коефіцієнтами). Отже, допоміжна задача має вигляд

$$f'(x) = Mx_3 + Mx_4,$$

$$\begin{cases} 4x_1 - x_2 - \lambda_2 - v_1 + x_3 = 10, \\ -x_1 + 4x_2 + \lambda_1 + \lambda_2 - v_2 + x_4 = 20, \\ x_2 + s_1 = 8, \\ x_1 + x_2 + s_2 = 10. \end{cases}$$

Складемо симплекс-таблицю 8.1.

Таблиця 8.1 – Вихідні дані

Базис	$C_{баз}$	C_j	0	0	0	0	0	0	M	M	0	0
		B	A_1	A_2	A_3	A_4	A_5	A_6	A_7	A_8	A_9	A_{10}
A_7	M	10	4	-1	0	-1	-1	0	1	0	0	0
A_8	M	20	-1	4	1	1	0	-1	0	1	0	0
A_9	0	8	0	1	0	0	0	0	0	0	1	0
A_{10}	0	10	1	1	0	0	0	0	0	0	0	1
	L_j	$30M$	$3M$	$3M$	M	0	$-M$	$-M$	M	M	0	0
	Δ_j		$3M$	$3M$	M	0	$-M$	$-M$	0	0	0	0

Оскільки в першу чергу необхідно мінімізувати нев'язання, виведемо з базису вектор $\overline{A}_7$ і введемо вектор $\overline{A}_1$ (тим самим змінна x_3 дорівнюватиме нулю, а змінна x_1 увійде до базису). Складемо нову симплекс-таблицю 8.2.

Таблиця 8.2 – Перша ітерація

Базис	$C_{баз}$	C_j	0	0	0	0	0	0	M	M	0	0
		B	A_1	A_2	A_3	A_4	A_5	A_6	A_7	A_8	A_9	A_{10}
A_1	0	10/4	1	-1/4	0	-1/4	-1/4	0	1/4	0	0	0
A_8	M	90/4	0	15/4	1	3/4	-1/4	-1	1/4	1	0	0
A_9	0	8	0	1	0	0	0	0	0	0	1	0
A_{10}	0	30/4	0	5/4	0	1/4	1/4	0	-1/4	0	0	1
	L_j	$90M/4$	0	$15M/4$	M	$3M/4$	$-M/4$	$-M$	$M/4$	M	0	0
	Δ_j		0	$15M/4$	M	$3M/4$	$-M/4$	$-M$	$3M/4$	0	0	0

Тепер виведемо з базису вектор $\overline{A}_8$ і введемо $\overline{A}_2$ у таблиці 8.3.

Таблиця 8.3 – Друга ітерація

Базис	$C_{баз}$	C_j	0	0	0	0	0	0	M	M	0	0
		B	A_1	A_2	A_3	A_4	A_5	A_6	A_7	A_8	A_9	A_{10}
A_1	0	4	1	0	1/15	-1/5	-4/15	-1/15	4/15	1/15	0	0
A_2	0	6	0	1	4/15	3/15	-1/15	-4/15	1/15	4/15	0	0
A_9	0	2	0	0	-4/15	-1/5	1/15	4/15	-1/15	-4/15	1	0
A_{10}	0	0	0	0	-1/3	0	1/3	1/3	-1/3	-1/3	0	1
	L_j	0	0	0	0	0	0	0	0	0	0	0
	Δ_j	0	0	0	0	0	0	0	$-M$	$-M$	0	0

Дістали припустимий базисний план вихідної задачі:

$$x_1=4;\ x_2=6;\ \lambda_1=0;\ \lambda_2=0;\ v_1=0;\ v_2=0;\ s_1=2;\ s_2=0.$$

Перевіримо виконання умов правила додаткової нетвердості (7.32) $\overline{X}^T\overline{V}=0$ і $\overline{\lambda}^T\overline{S}=0$:

$$x_1{*}v_1=0;\ x_2{*}v_2=0;\ \lambda_1{*}s_1=0;\ \lambda_2{*}s_2=0.$$

Оскільки умови додаткової нетвердості виконуються, отриманий план є оптимальним. Оптимальне значення цільової функції

$$f(x_1, x_2)= 10{*}4+20{*}6+4{*}6-2{*}4^2-2{*}6^2= 80.$$

Контрольні запитання

1. За яких умов оптимізаційну задачу можна віднести до класу нелінійних?
2. Наведіть приклад економічної моделі, що збігається до задачі нелінійного програмування.
3. Перелічіть основні труднощі, що виникають у процесі розв’язання задачі нелінійного програмування.
4. Який зміст вкладають в поняття «умовна оптимізація»?
5. Для чого призначений метод множників Лагранжа та у чому він полягає?
6. Яку точку множини розв’язків називають стаціонарною?
7. Які принципові етапи належать до градієнтних методів?

8. Поясніть, для розв’язання яких задач призначений метод найскорішого спуску й метод дроблення кроку?

9. Дайте визначення опуклої (увігнутої) функції.

10. Сформулюйте достатню умову опуклості (увігнутості) функції.

11. У чому полягає специфіка задач опуклого програмування?

12. Дайте визначення сідлової точки. Наведіть приклад функції, що має сідлову точку.

13. Сформулюйте необхідну й достатню умови теореми Куна-Таккера. Яке значення вони мають для розв’язання задач нелінійного програмування?

14. У чому полягає умова регулярності Слейтера? Поясніть її зміст.

15. Наведіть приклад пари двоїстих задач нелінійного програмування.

16. Які властивості пари нелінійних двоїстих задач можна застосувати для їх розв’язання?

ЗМ 2 Аналіз та управління ризиком в економіці

ТЕМА 9
ПОНЯТТЯ ЕКОНОМІЧНОГО РИЗИКУ. КЛАСИФІКАЦІЯ, МЕТОДИ ОЦІНКИ І УПРАВЛІННЯ

9.1 Поняття економічного ризику

В умовах ринкових відносин проблема оцінки й урахування економічного ризику здобуває самостійне значення як важлива частина менеджменту, теорії й практики управління. Вільна взаємодія ринкових суб'єктів і ринкова конкуренція неминуче підвищують невизначеність і комерційний ризик.

Більшість управлінських рішень приймається в умовах ризику, що зумовлено рядом факторів: відсутністю повної інформації, наявністю протиборчих тенденцій, елементами випадковості й багатьом іншим.

Посилення ризику - це зворотний бік свободи підприємництва. Щоб вижити в умовах ринкових відносин, підприємцеві потрібно зважуватися на впровадження технічних нововведень і на сміливі, нетривіальні дії, а це посилює ризик. Для будь-якого бізнесу важливим є не запобігання ризику взагалі, а передбачення й зниження його до мінімального рівня.

Розгляд фінансових ризиків вимагає залучення апарата фінансової математики. У науці про фінанси важливою є оцінка прибутку й ризику фінансової операції. Особливістю фінансових ризиків є ймовірність настання збитку в результаті здійснення таких операцій, які за своєю природою є ризикованими.

Проблема кількісної і якісної оцінки економічних ризиків і управління ризиками є актуальною.

На сьогодні немає однозначного розуміння сутності ризику. Ризик - це складне явище, що має множину незбіжних, а іноді протилежних реальних основ. Це зумовлює можливість існування кількох визначень ризику з різних точок зору. Аналіз численних визначень ризику дозволяє виявити основні моменти, характерні для ризикової ситуації, такі як:

- випадковий характер події, що визначає, який з можливих наслідків реалізується на практиці (наявність невизначеності);
- наявність альтернативних рішень;
- відомі або можна визначити імовірності наслідків і очікувані результати;
- імовірність виникнення збитків;
- імовірність одержання додаткового прибутку.

Зупинимося на наступному визначенні ризику. Це діяльність, пов'язана з подоланням невизначеності в ситуації неминучого вибору, у процесі якої є можливість кількісно і якісно оцінити імовірність досягнення передбачуваного результату, невдачі й відхилення від цілі.

Слід зазначити, що різниця між ризиком і невизначеністю належить до способу завдання інформації й визначається наявністю (у випадку ризику) або відсутністю (при невизначеності) імовірнісних характеристик неконтрольованих змінних. У такому смислі ці терміни вживають в математичній теорії дослідження операцій, де розрізняють задачі прийняття рішень при ризику й відповідно в умовах невизначеності.

У зв'язку з тим, що ризик є специфічною діяльністю в умовах невизначеності й у ситуації обов'язкового вибору, то він також являє собою діалектичну єдність об'єктивного й суб'єктивного. Отже, ризик завжди пов'язаний з вибором певних альтернатив і розрахунком імовірності їх результату - у цьому проявляється його суб'єктивна сторона. Разом з тим, величина ризику не тільки суб'єктивна, але й об'єктивна, оскільки вона є формою якісно-кількісного виразу реально існуючої невизначеності.

Для розуміння природи підприємницького ризику фундаментальне значення має зв'язок ризику й прибутку. Підприємець проявляє готовність йти на ризик в умовах невизначеності, оскільки поряд з ризиком втрат існує можливість додаткових прибутків. Хоча ясно, що одержання прибутку підприємцеві не гарантовано, винагородою за витрачений ним час, зусилля й спроможності можуть виявитися як прибуток, так і втрати.

На рисунку 9.1 показано залежність прибутку від ризику. Можна вибрати рішення, що містить менше ризику (r_1=0), але при цьому меншим буде й одержуваний прибуток ($П_1$), а при найвищому ризику r_3 прибуток має найбільше значення, що дорівнює $П_3$.

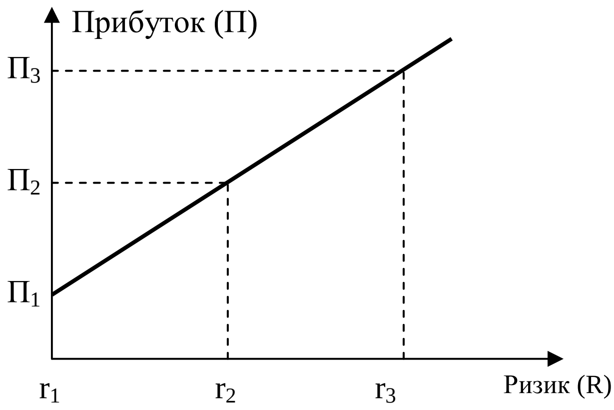


Рисунок 9.1 – Залежність прибутку від ризику

Ризик підприємця, як правило, орієнтований на одержання значущих результатів нетрадиційними методами. Його можна розглядати як невід'ємний елемент підприємницької діяльності. Так, в економічній боротьбі з конкурентами-виробниками за покупця, підприємницька організація змушена продавати свою продукцію в кредит (з ризиком неповернення грошових сум у строк), при наявності тимчасово вільних коштів розміщати їх у вигляді депозитних внесків або цінних паперів (з ризиком одержання недостатнього процентного доходу в порівнянні з темпами інфляції), при веденні комерційних операцій експортно-імпортного характеру зіштовхуватися з необхідністю оперувати різними національними валютами (з ризиком втрат від несприятливої кон'юнктури курсів валют) та ін.

Наявність ризику передбачає необхідність вибору одного з можливих варіантів рішень, у зв'язку з чим необхідний аналіз всіх можливих альтернатив і вибір найрентабельніших і найменш ризикованих з них. Залежно від конкретного змісту ситуації ризику альтернативність має різний ступінь складності. Тому в складних економічних ситуаціях для вибору оптимального рішення використовують спеціальні методи аналізу.

Отже, можна зробити висновок, що, незважаючи на значний потенціал втрат, що несе в собі ризик, він є й джерелом можливого прибутку. Тому основним завданням підприємця є не відмова від ризику взагалі, а вибір рішень, пов'язаних з ризиком, на основі об'єктивних критеріїв.

9.2 Класифікація ризиків

Класифікація ризиків є досить складною проблемою, що зумовлена їх різноманіттям.

За характером наслідків ризики підрозділяють на чисті й спекулятивні.

Особливість чистих ризиків (їх іноді називають статистичними або простими) полягає в тому, що вони практично завжди несуть у собі втрати для підприємницької діяльності. Їх причинами можуть бути стихійні лиха, нещасні випадки, недієздатність керівників фірм та ін.

Спекулятивні ризики, які називають також динамічними або комерційними, несуть у собі або втрати, або додатковий прибуток для підприємця. Їх причинами можуть бути зміна курсів валют, зміна кон'юнктури ринку, зміна умов інвестування та ін.

За сферою виникнення, в основу якої покладені сфери діяльності, розрізняють наступні види ризиків:

виробничий ризик, пов'язаний з невиконанням підприємством своїх планів і зобов'язань з виробництва продукції, товарів, послуг, інших видів виробничої діяльності в результаті впливу як зовнішнього середовища, так і внутрішніх факторів;

комерційний ризик - це ризик втрат у процесі фінансово-господарської діяльності; його причинами можуть бути зниження обсягів реалізації, непередбачене зниження обсягів закупівель, підвищення закупівельної ціни товару, підвищення витрат обігу, втрати товару в процесі обігу та ін.;

фінансовий ризик виникає у зв'язку з неможливістю виконання фірмою своїх фінансових зобов'язань, їх причинами є зміна купівельної спроможності грошей, нездійснення платежів, зміна валютних курсів та ін.

Залежно від причини виникнення ризиків, їх поділяють на наступні категорії:

природні ризики - це ризики пов'язані із проявом стихійних сил природи;

екологічні ризики пов'язані з настанням громадянської відповідальності за завдання збитків навколишньому середовищу;

політичні ризики - це можливість виникнення збитків або скорочення розмірів прибутку, що є наслідком державної політики;

транспортні ризики пов'язані з перевезеннями вантажів різними видами транспорту;

майнові ризики - це ризики від втрати майна підприємця з причин від нього не залежних;

торгові ризики залежать від збитків через затримку платежів, не постачання товару, відмови від платежу та ін.

Велика група ризиків пов'язана з купівельною спроможністю грошей. До неї належать:

інфляційні ризики, які зумовлені знецінюванням реальної купівельної спроможності грошей, при цьому підприємець зазнає реальних втрат;

дефляційний ризик пов'язаний з тим, що при зростанні дефляції падає рівень цін і, отже, знижуються доходи;

валютні ризики пов'язані із зміною валютних курсів, вони належать до спекулятивних ризиків, тому, при втратах однієї із сторін у результаті зміни валютних курсів, інша сторона, як правило, отримує додатковий прибуток і навпаки;

ризик ліквідності пов'язаний з втратами при реалізації цінних паперів або інших товарів через зміну оцінки їх якості й споживчої вартості.

Інвестиційні ризики пов'язані з можливістю недоодержання або втрати прибутку в ході реалізації інвестиційних проектів, вони містять у собі наступні підвиди ризиків:

ризик упущеної вигоди полягає в тому, що виникає фінансовий збиток у результаті нездійснення певного заходу;

ризик зниження прибутковості пов'язаний із зменшенням розміру відсотків і дивідендів за портфельними інвестиціями; його поділяють на процентний ризик, що виникає в результаті перевищення процентних ставок, виплачуваних за притягнутими коштами, над ставками за наданими кредитами, і кредитний ризик, що виникає у випадку несплати позичальником основного боргу й відсотків, що належать кредитору;

біржові ризики являють собою небезпеку втрат від біржових операцій;

селективні ризики виникають через неправильне формування видів вкладення капіталів, виду цінних паперів для інвестування;

ризик банкрутства пов'язаний з повною втратою підприємцем власного капіталу через його неправильне вкладення.

Розглянемо такий критерій як припустима границя ризику.

Під **припустимим ризиком** розуміють рівень ризику в межах його середнього рівня, тобто середнього щодо інших видів діяльності й інших господарчих суб'єктів. Якщо позначити через R середній рівень ризику, а через R_D - рівень припустимого ризику, то повинна виконуватись нерівність

$$R_{\mathrm{D}}<R. \tag{9.1}$$

Під **критичним ризиком** $R_{кр}$ розуміють ризик, рівень якого вище за середній, але в межах максимально припустимих значень ризику $\mathrm{R}_{\max}$ прийнятих у даній економічній системі для певних видів діяльності, тобто

$$R_{\mathrm{D}} < R_{\text{кр}} < R_{\max}. \tag{9.2}$$

Катастрофічний ризик $R_{\text{кат}}$ - це такий ризик, що перевищує максимальну границю ризику $R_{\max}$, яка склалася в даній економічній системі, і для якого виконується умова

$$R_{\text{кат}} > R_{\max}. \tag{9.3}$$

Ризикована ситуація пов'язана з статистичними процесами і її супроводжують три умови: наявність невизначеності, необхідність вибору альтернативи й можливість якісної й кількісної оцінки імовірності здійснення того або іншого варіанта.

Розглядаючи діяльність певної економічної системи, потрібно враховувати, що вона завжди сполучена з невизначеністю. Наявність невизначеності зумовлює виникнення ризиків, без урахування яких неможливий ефективний розвиток економіки.

Невизначеність - це неповне або неточне уявлення про значення параметрів у майбутньому, у тому числі про умови реалізації рішення й пов'язані з ними витрати і результати. Невизначеність, пов'язана з можливістю виникнення в ході реалізації рішення несприятливих ситуацій і їх наслідків, характеризується поняттям ризик.

З погляду імовірності подій невизначеність можна поділити на три види: повна невизначеність, повна визначеність, часткова невизначеність.

Повна невизначеність характеризується близькою до нуля прогностичністю P_t настання події, що математично виражається співвідношенням

$$\lim_{t \to t_k} P_t = 0, \tag{9.4}$$

де t - час;

t_k - кінцевий час прогнозування події.

Повній визначеності відповідає близька до одиниці прогнозованість подій, тобто

$$\lim_{t \to t_k} P_t = 1. \tag{9.5}$$

Це можливо, насамперед, у тих випадках, коли при вирішенні задачі в умовах невизначеності з певною імовірністю можна визначити оптимальне рішення й з заздалегідь відомою імовірністю (зазвичай рівною 0,9 -0,99) знайти довірчий прогнозований інтервал.

Часткова невизначеність відповідає таким подіям, прогностичність яких лежить у межах від 0 до 1, що визначається нерівністю

$$0 < \lim_{t \to t_k} P_t < 1. \tag{9.6}$$

В умовах об'єктивного існування ризику й пов'язаних з ним фінансових, моральних та інших втрат виникає потреба в певному механізмі, що дозволив би найкращим з можливих способів з погляду поставлених підприємцем або фірмою цілей ураховувати ризик при прийнятті й реалізації господарської діяльності.

9.3 Методи оцінки економічних ризиків

Кількісну оцінку економічного ризику здійснюють для вибору із сукупності альтернативних рішень оптимального рішення. Тут під оптимальним розуміють рішення, що забезпечує найбільшу імовірність найкращого результату при найменших витратах і втратах відповідно до задач мінімізації й програмування ризику. Для оцінки економічного ризику використовують математичні методи і моделі.

Застосування економіко-математичних методів дозволяє провести якісний і кількісний аналіз економічних явищ, дати кількісну оцінку значення ризику й ринкової невизначеності та вибрати оптимальне рішення.

Як математичний апарат прийняття рішень в умовах невизначеності й ризику використовують методи математичного програмування, теорії ігор, теорії імовірностей і математичної статистики.

Теорія ігор - це теорія математичних моделей прийняття оптимальних рішень в умовах невизначеності, протилежних інтересів різних сторін, конфлікту. Матричні ігри можуть служити математичними моделями багатьох найпростіших конфліктних ситуацій з галузі економіки. Зокрема, теорія ігор застосовується в питаннях боротьби фірм за ринки, у явищах олігополії, у плануванні рекламних компаній, при формуванні цін на конкурентних ринках, у біржовій грі та ін. З позицій теорії ігор можна розглядати питання централізації й децентралізації керування виробництвом, оптимальне планування за кількома показниками, планування в умовах невизначеності, породжуваної, наприклад, технічним прогресом, подолання відомчих протидій та ін.

Методи прийняття рішень в умовах ризику також розробляють й обґрунтовують в рамках так званої теорії статистичних рішень. Суть статистичного методу, як уже вказувалося, полягає в тому, що аналізується статистика збитків і прибутків, що мали місце на даному або аналогічному підприємстві (економічна ситуація), встановлюється величина й частість одержання того або іншого економічного результату й складається найбільш імовірний прогноз на майбутнє. Недоліком статистичного підходу до виміру ризику є той факт, що він заснований на наявних статистичних даних минулих періодів, у той час як оцінка ризику стосується майбутніх подій. Це знижує цінність даного підходу в умовах мінливої економічної обстановки. У той самий час перевагою даного підходу до виміру ризику є його об'єктивність.

Економіко-математичні задачі, ціль яких полягає в знаходженні найкращого (оптимального) з погляду певного критерію (або критеріїв) варіанта використання наявних ресурсів (праці, капіталу, та ін.), називаються оптимізаційними. Оптимізаційні задачі вирішують за допомогою оптимізаційних моделей методами математичного програмування. Необхідною умовою використання оптимального підходу до планування й управління (принципу оптимальності) є гнучкість, альтернативність виробничо-господарських ситуацій, в умовах яких доводиться приймати планово-управлінські рішення. Саме такі ситуації, як правило, і складають повсякденну практику господарюючого суб'єкта (вибір оптимальних асортиментів виробничої програми, прикріплення до постачальників, складання портфеля цінних паперів, вкладення інвестицій до оптимального проекту, маршрутизація, розкрій матеріалів та ін.).

9.4 Управління ризиками

При управлінні ризиком у першу чергу оцінюють ступінь ризику, тобто роблять кількісний аналіз, що передбачає числове визначення окремих ризиків і ризику рішення в цілому. На цьому етапі визначають числові значення імовірності настання ризикових подій і їх наслідків, а також припустимий у даній конкретній обстановці рівень ризику.

В результаті проведення аналізу ризику одержують картину можливих ризикових подій, імовірність їх настання й наслідків. Після порівняння отриманих значень ризиків із гранично припустимими виробляють стратегію управління ризиком, і на цій основі - заходи запобігання й зменшення ризику.

Після вибору певного набору заходів щодо усунення й мінімізації ризику, треба ухвалити рішення щодо достатності обраних заходів. У випадку достатності здійснюється реалізація проекту (прийняття частини ризику, що залишилася), у протилежному випадку доцільно відмовитися від реалізації проекту (уникнути ризику).

Основні величини оцінки економічного ризику пов'язані з виграшем або програшем, із втратою або прибутком у результаті господарської або підприємницької діяльності. Теоретичний спосіб економічної оцінки ризику ґрунтується на двох параметрах: розмірі можливих втрат і їх імовірності, а це означає, що потрібно визначити кількісну величину обох характеристик, після чого ризики стають порівняними один з одним.

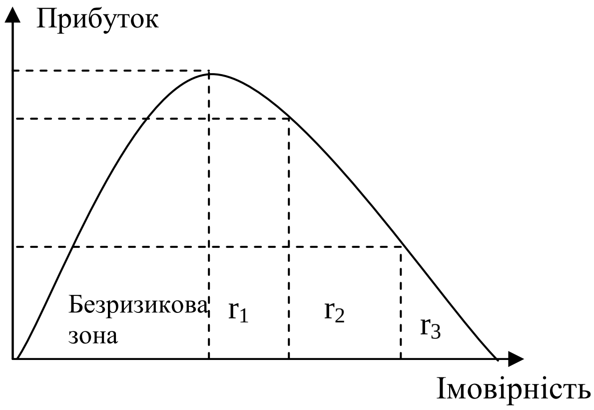


Рисунок 9.2 – Схема імовірності отримання прибутку

Розглянемо схему імовірності одержання прибутку. На рисунку 9.2 представлена крива ризику й зони ризику: припустима r_1, підвищеного ризику r_2, і критичного ризику r_3 залежно від величини прибутку, а також безризикова зона, у якій не очікується непередбачених втрат.

Крива імовірності ризику є множиною точок, які встановлюються для кожного значення величини можливого прибутку й відповідної імовірності її виникнення. Крива ризику, будучи початковою стадією аналізу ризикової ситуації, наочно показує підприємцеві ефективні зони ризику й дає кількісну оцінку ризикового прибутку, що бажає одержати підприємець.

Управління ризиками ґрунтується на об'єктивних знаннях про їх характер, прогнозування й своєчасну оцінку негативних факторів, що впливають на успіх реалізації прийнятого рішення.

Незалежно від причин виникнення економічного ризику природним є бажання кожного суб'єкта зменшити можливі втрати, пов'язані з реалізацією економічного ризику.

Метою розробки будь-якої моделі управління ризиком є забезпечення успішного функціонування ризикового проекту. Цієї мети можна досягнути за рахунок рішення наступних основних задач - виявлення можливих економічних ризиків і зниження фінансових втрат, пов'язаних з економічними ризиками.

Розглянемо модель управління ризиком, що є послідовністю дій, які дозволяють дотримувати розумних сполучень ризиків і вигід розглянутих проектів.

Учасники проектів включають найрізноманітніших «гравців»: виробників, транспортні організації, споживачів, банки, торговельні й промислові підприємства. При реалізації проекту всі вони мають свої економічні задачі, інтереси й сформовані стратегії, а відповідно і моделі управління тими видами ризику, які мають до них відношення. Вибір варіантів управління може бути різним, так само як формовані портфелі інструментів для управління ризиком, які розвиваються згодом, пристосовуючись до мінливих ринкових умов. Однак технологія управління ризиками є незмінною й включає наступні послідовно виконувані елементи:

- встановлення ризиків (виявлення джерел і типів ризику);
- оцінку (вимірювання) ризиків, аналіз факторів і умов, що впливають на імовірність ризику, розміри втрат і збитків, а також граничні (нормативні) рівні ризиків;
- вибір способів і визначення засобів для скорочення й утримання ризиків;
- ситуаційний контроль за ризиками, порівняння із припустимим (нормативним) рівнем ризиків, проведення заходів щодо вирішення ризиків, їх коректування з урахуванням складної ситуації; покриття збитків і ліквідацію інших негативних наслідків прояви ризиків; нагромадження й обробку ретроспективної інформації про ризикові ситуації й наслідки прояви ризиків, вироблення рекомендацій для урахування отриманого досвіду в майбутньому.

Класифікація ризиків є одним з етапів аналізу ризиків, що дозволяє надалі здійснювати ідентифікацію й оцінку ризику, а також розробляти методи управління ними. Класифікація полягає в розподілі ризиків за групами на основі класифікаційних критеріїв. Залежно від цілей дослідження на практиці й у науковій літературі використовують різні класифікації. Зустрічаються класифікації, засновані на розходженнях видів діяльності (фінансові ризики, виробничі ризики, ризики матеріально-технічного постачання та ін.). Часто ризики класифікують за областями їх прояву (політичні ризики, соціальні ризики, природні ризики та ін.).

Групи ризику виділяють і за іншими специфічними ознаками.

В основі найпоширенішої класифікації лежить виділення чистих і спекулятивних ризиків. Чисті ризики мають відносно постійний характер. Для їх аналізу й оцінки використовують методи математичної статистики й теорії імовірностей, оскільки їх прояв, як правило, стабільний в часі або відрізняється певною закономірністю. Стабільний і сталий характер динаміки основних показників чистих ризиків дозволяє називати їх також статистичними ризиками.

На відміну від чистих, спекулятивні ризики повною мірою визначаються управлінським рішенням. Нерідко спекулятивні ризики мають невизначений характер прояву, їх аналітичні оцінки змінюються з часом.

Класифікація й ідентифікація ризиків необхідна для їх своєчасної оцінки, прогнозування негативних факторів при реалізації інноваційних процесів. Знання про характер ризиків, їх ідентифікація за видами і базовими ознаками дозволяє розробляти заходи щодо зниження ризиків в операційній, інвестиційній і фінансовій діяльності.

Фактори ризиків визначають на основі аналізу політичної, економічної й фінансово-кредитної політики й відіграють першорядну роль у прийнятті рішень. Фактори ризиків - одна із самих складних частин і в той самий час один з ключових напрямків роботи з управління ризиком. Проводити факторний аналіз набагато складніше, ніж будь-який інший, оскільки ті самі фактори роблять у різних умовах неоднаковий вплив на ринок або можуть з вирішальних стать абсолютно незначущими.

Розгляд повного переліку ситуацій, що виникають при реалізації певного процесу, на практиці не тільки неможливий, але й економічно недоцільний. У таких випадках говорять про існування фактору невизначеності, і під невизначеністю в цьому випадку розуміють неможливість повного й вичерпного аналізу всіх факторів, що впливають на результат конкретних фінансових вкладень.

Доцільність ухвалення конкретного рішення у розглянутій економічній ситуації, що споконвічно містить певний ступінь ризику, можна виявити шляхом кількісної оцінки ризику. Ця оцінка особливо важлива, коли існує можливість вибору конкретного рішення з сукупності альтернативних варіантів.

У процесі управління ризиками необхідно вибрати стратегію, що дозволяє зменшити ступінь ризику. Математичний апарат для вибору стратегії в конфліктних ситуаціях дає теорія ігор. Аналіз ситуацій за допомогою методів теорії ігор дозволяє розглянути всі можливі альтернативи як своїх дій, так і стратегії партнерів і конкурентів. За допомогою теорії ігор можна вирішувати багато економічних задач, що пов'язані з вибором найкращого рішення. Звідси видно, що ризик має математично виражену імовірність настання збитку, ця імовірність спирається на статистичні дані й може бути розрахованою з досить високим ступенем точності.

Для аналізу та управління системою ризику можна свикористатися алгоритмом ризик-менеджменту, зміст і задачі якого подані на рисунку 9.3.

На першому етапі відносно нескладно визначити перелік заходів щодо зниження ризику, які необхідно здійснити у першу чергу. Однак потім неминуче встає питання про способи подальшого зниження рівня ризику. І в цьому випадку проведення додаткових попереджувальних заходів уже не здається очевидним, оскільки всі вони пов'язані з певними витратами. Аналіз розглянутих вище характеристик різних методів зниження ризику дозволяє зробити висновок про те, що будь-який захід, спрямований на зниження ризику, як правило, має свою ціну.

Застосування кожного з методів управління ризиком призводить до перерозподілу поточних і очікуваних фінансових потоків. Наприклад, при страхуванні частина власних коштів відволікається на сплату страхових внесків, при страхуванні за допомогою розподілу ризику (передачі частини ризику) доводиться йти на відмову від частини прибутків, у результаті чого відбувається недоінвестування проекту і втрата прибутку.

При хеджуванні за допомогою опціонів платою за зниження ризику є опціонна премія. При диверсифікованості зменшення ризику призводить до зниження очікуваної віддачі (дивідендів, прибутку та ін.). Резервування пов'язане з витратами на утримання резервних фондів, що веде до відволікання оборотних коштів і в остаточному підсумку призводить до зниження прибутку. Зменшення ризику за допомогою методу лімітування призводить до зниження віддачі як наслідку прийнятих обмежень.

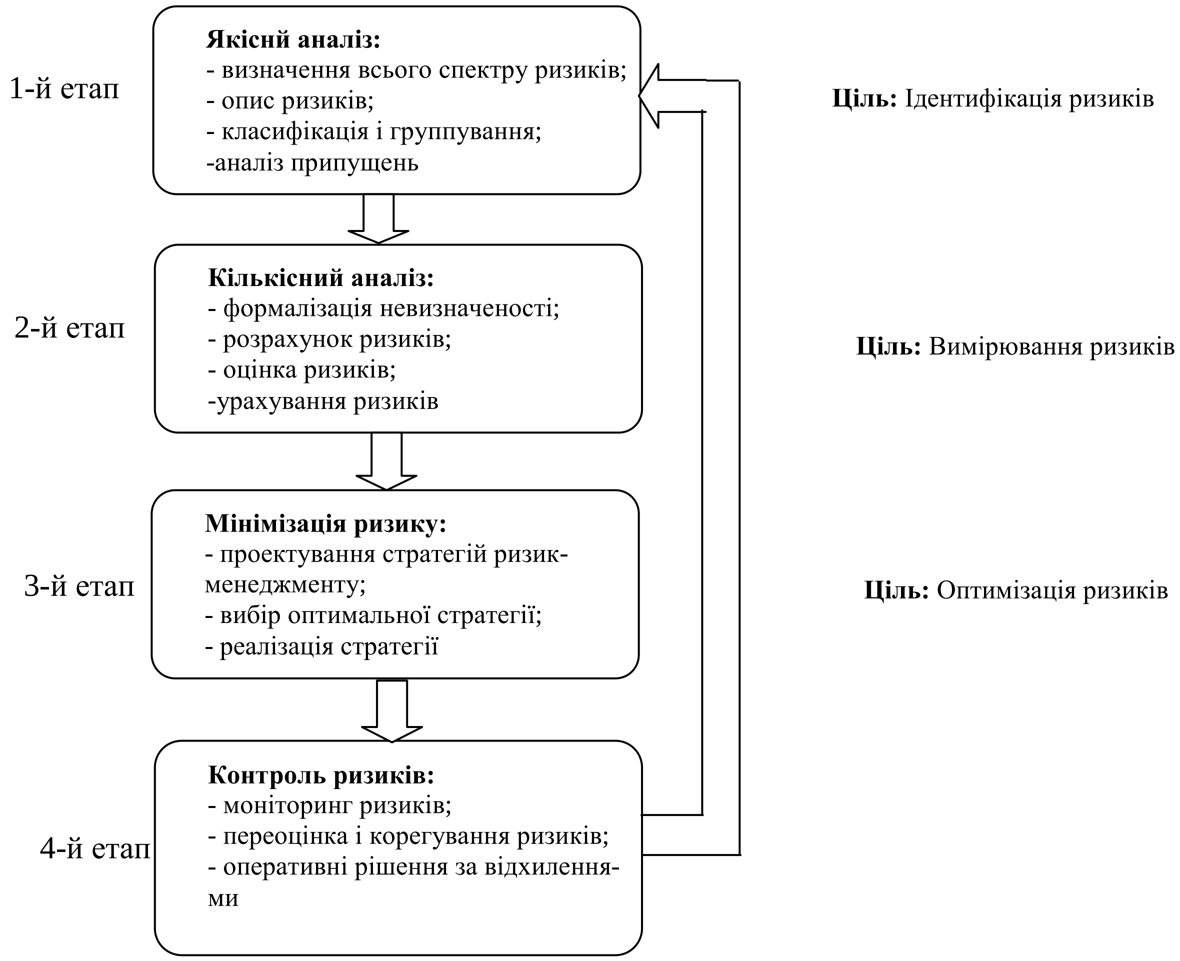


Рисунок 9.3 - Алгоритм управління ризиком

З іншого боку виникає очікуваний у майбутньому приплив коштів у вигляді компенсації за зменшення ризику. Видно, що кожний з методів зниження ризику відрізняється ступенем впливу на ризик у конкретній ситуації, а також необхідними витратами на їх реалізацію.

Як критерій економічної ефективності застосування методів управління можна використовувати оцінку їх впливу на зміну вартості проекту (підприємства або ситуації в цілому), розраховану на початок і кінець фінансового періоду. Необхідно оцінити динаміку ризику й віддачі, порівняти отримані результати з цілями й ступенем їх досягнення й зробити висновок про економічну доцільність розглянутих заходів.

Існують різні шляхи підходу до вибору способів попередження й зниження ризиків, сутність яких відображає наведена в таблиці 9.1 спрощена схема рішень із зниження ступеня ризику.

Градація рівнів ризику й можливих втрат, представлена в таблиці, є приблизною (умовною). Подібна градація наведена при розгляданні зон ризику з конкретними значеннями імовірностей P_j. У таблиці також відсутні такі методи зниження ризику як диверсифікованість, резервування, лімітування та ін.

Таблиця 9.1 - Рішення із зниження ризику

<table>
<tr><th>Імовірність втрат (рівень втрат)</th><th>Близька до нуля</th><th>Низька</th><th>Невелика</th><th>Середня</th><th>Велика</th><th>Близька до одиниці</th></tr>
<tr><td>Незначний (від 0 до P_j)</td><td colspan="4">Прийняття ризику</td><td colspan="2">Прийняття ризику або створення резервів, запасів</td></tr>
<tr><td>Малий (від P_1 до P_2)</td><td colspan="6">Створення резервів, запасів</td></tr>
<tr><td>Припустимий (від P_2 до P_3)</td><td colspan="2">Створення резервів, запасів</td><td colspan="3">Зовнішнє страхування або (і) розподіл ризику</td><td>Запобігання ризику</td></tr>
<tr><td>Середній (від P_3 до P_4)</td><td colspan="4">Зовнішнє страхування або (і) розподіл ризику</td><td colspan="2">Запобігання ризику</td></tr>
<tr><td>Великий (від P_4 до P_5)</td><td colspan="3">Зовнішнє страхування або (і) розподіл ризику</td><td colspan="3">Запобігання ризику</td></tr>
<tr><td>Катастрофічний (більший за P_5)</td><td colspan="2">Зовнішнє страхування або (і) розподіл ризику</td><td colspan="4">Запобігання ризику</td></tr>
</table>

З таблиці видно, що при незначних втратах, незалежно від імовірності їх виникнення, рекомендується прийняття ризику. При малих втратах і при припустимих втратах у випадку низької імовірності їх виникнення доцільно застосовувати резервування. При подальшому збільшенні втрат залежно від імовірності їх виникнення рекомендується зовнішнє страхування або розподіл ризику й при збільшенні втрат - запобігання ризику.

Відзначимо, що починаючи з певного рівня, витрати на пряме зниження ризику зростають скоріше, чим знижується сам ризик.

Застосування лише одного методу зниження ризику є недоцільним, тільки комбінація методів у різних сполученнях, дозволить досягти оптимального співвідношення між рівнем досягнутого зниження ризику й необхідними для цього додатковими витратами.

При розробці моделі управління ризиками основною метою є зменшення можливого ризику, зниження фінансового збитку.

Розглядаючи певний ризиковий економічний проект, ставлять цілі, завдання й принципи управління ризиками, дають попередню класифікацію збитків, оцінюють можливі збитки й вибирають основні методи управління ризиками. Цю частину аналізу треба вважати попередньою і тут також потрібно проаналізувати всю інформацію з подібних проектів.

З огляду на загальну схему процесу управління ризиком і спираючись на попередній аналіз, подальша розробка моделі відбувається в наступній послідовності:

- з метою вибору можливих методів управління ризиками визначають припустимі значення ризиків за кожним з методів, і у зв'язку з цим уточнюють стратегію фірми;
- використовуючи дані попереднього пункту, роблять відбір ризиків, у результаті якого відмовляються від тих ризиків, які є неприйнятними для даного проекту;
- з метою зниження збитку від обраних ризиків проводять дослідження з вироблення попереджуючих заходів;
- з урахуванням цих заходів ризики аналізують з метою виявлення можливостей застосування до них намічених методів управління ризиками;
- складають остаточний план попереджуючих заходів, знаходять ступінь оцінки ризиків, величину можливого збитку, імовірність його настання. Визначають довірчі границі збитку, тобто вказують значення максимального й мінімального можливого, найбільш імовірного. Рекомендується оптимальний варіант проекту;
- аналізують ефективність проекту з метою оцінки його розробки й впровадження.

Управління ризиком в останні роки виділяється в самостійний вид професійної діяльності, у якій задіяні інститути фахівців, страхові компанії, а також фінансові менеджери, менеджери з ризику, фахівці із страхування.

Ще раз відзначимо, що ризик властивий підприємництву і є невід'ємною частиною його економічного життя. Тому відповідні знання про ризик, фактори, що впливають на його значущість, і способи зниження його негативного впливу, є необхідним інструментом, що зумовлює ефективність підприємницької діяльності.

Контрольні запитання

1. У чому полягає роль ризику підприємництва у ринкових умовах?
2. Якими факторами зумовлене виникнення ризикової ситуації?
3. Поясніть, у чому полягає зв'язок ризику й прибутку?
4. Чим зумовлена складність класифікації ризиків?
5. Які ризики належать до чистих ризиків і які до спекулятивних?
6. Як класифікують ризики за сферою виникнення?
7. Як відрізняють припустимий, критичний і катастрофічний ризики?
8. Охарактеризуйте поняття «невизначеність».
9. Який математичний апарат використовують для якісної та кількісної оцінки ризику?
10. На яких двох параметрах ґрунтується теоретичний спосіб економічної оцінки ризику?

ТЕМА 10
ПРИЙНЯТТЯ РІШЕНЬ В УМОВАХ НЕВИЗНАЧЕНОСТІ ТА РИЗИКУ

Наявність невизначеностей значно ускладнює процес вибору оптимальних рішень і може призвести до непередбачених результатів. У цілому ряді економічних задач доводиться аналізувати ситуації, у яких необхідно приймати рішення в умовах невизначеності, тобто, наприклад, виникають ситуації, у яких зіштовхуються інтереси двох або більше конкуруючих сторін, кожна з яких переслідує свою мету, причому результат будь-якого заходу кожної із сторін залежить від того, які дії почне супротивник. Це є характерним особливо в умовах ринкової економіки. Такі ситуації називають конфліктними. Науково обґрунтовані методи вирішення задач із конфліктними ситуаціями дає теорія ігор.

10.1 Система невизначеностей

Невизначеність, що має місце при будь-яких видах підприємницької діяльності, зумовлена тим, що економічна система в процесі функціонування зазнає впливу від цілого ряду факторів.

За часом виникнення невизначеності поділяють на **ретроспективні**, **поточні** та **перспективні**. Необхідність урахування фактору часу при оцінці економічної ефективності прийнятих рішень зумовлена тим, що як ефект, так і витрати можуть бути розподілені в часі. Рівні за величиною витрати, по-різному розподілені в часі, забезпечують неоднаковий корисний результат того або іншого виду.

За факторами виникнення невизначеності поділяють на **економічні** (комерційні) і **політичні**. Економічні невизначеності зумовлені несприятливими змінами в економіці підприємства або в економіці країни, до них належать: невизначеність ринкового попиту, слабка передбачуваність ринкових цін, невизначеність ринкової пропозиції, недостатність інформації про дії конкурентів та ін. Політичні невизначеності зумовлені зміною політичної обстановки, що впливає на підприємницьку діяльність. Ці види невизначеності пов'язані між собою, і часто на практиці їх важко розділити.

Природна невизначеність описується сукупністю факторів, серед яких можуть бути кліматичні, погодні умови, різного роду перешкоди (атмосферні, електромагнітні та ін.).

Наступним видом невизначеності є **невизначеність зовнішнього середовища**. При економічному аналізі підприємницької діяльності вводять поняття зовнішнього й внутрішнього середовища. Внутрішнє середовище включає фактори, зумовлені діяльністю самого підприємця і його контактів. Зовнішнє середовище представлене факторами, які не пов'язані безпосередньо з діяльністю підприємця й мають більш широкий соціальний, демографічний, політичний і інший характер.

Особливий вид невизначеності має місце при наявності **конфліктних ситуацій**, якими можуть бути стратегія й тактика осіб, що беруть участь у тім або іншому конкурсі, дії конкурентів, цінова політика олігополістів та ін.

Відособлену групу складають задачі, у яких розглядаються проблеми незбіжних інтересів і багатокритеріального вибору оптимальних рішень в умовах невизначеності.

Розглядаючи невизначеність, що є найхарактернішою причиною ризику в економічній діяльності, необхідно зазначити, що виділення й вивчення її у процесі економічного, комерційного, управлінського, фінансового та інших видів діяльності є вкрай необхідним, оскільки при цьому відображається практична ситуація, коли немає можливості здійснювати перелічені види діяльності в умовах, які не можуть бути однозначно визначені.

10.2 Основні поняття теорії ігор

Теорія ігор - це теорія математичних моделей прийняття оптимальних рішень в умовах невизначеності, протилежних інтересів різних сторін конфлікту. Помітимо, що при цьому конфлікт не обов'язково повинен бути антагоністичним, як конфлікт можна розглядати будь-яку розбіжність. Важливим класом ігор є ігри з природою, у яких ризик пов'язаний з сукупністю невизначених факторів навколишнього середовища, іменованих «природа». Термін «природа» характеризує якусь об'єктивну дійсність, що не слід розуміти буквально, хоча цілком можуть зустрічатися ситуації, у яких гравцем дійсно може виступати природа (наприклад, обставини, пов'язані з погодними умовами або з природними стихійними силами).

Одним з основних питань у задачах з колективним вибором рішень є питання про визначення оптимальності, тобто питання, які рішення можна визнавати найкращими в ситуації оптимізації за кількома критеріями, що відображає різні інтереси. Багато методів вирішення проблем теорії ігор ґрунтуються на зведенні їх до задач математичного програмування.

Термін **гра** застосовують для позначення сукупності правил і угод, якими керуються суб'єкти, поводження яких ми вивчаємо. Кожний такий суб'єкт k, де $k=\overline{1,K}$, або **гравець**, характеризується наявністю індивідуальної системи цільових настанов і **стратегій** $s_1^k, s_2^k, \ldots, s_{m_k}^k$, тобто можливих варіантів дій у грі.

Досить розповсюджений спосіб математичного опису гри заснований на завданні функцій $f_k(s_{i1}^1, s_{i2}^2, \ldots, s_{ik}^k, \ldots, s_{iK}^K)$, кожна з яких визначає результат (платіж, виграш), одержуваний k-м гравцем залежно від набору стратегій $S=(s_{i1}^1, s_{i2}^2, \ldots, s_{ik}^k, \ldots, s_{iK}^K)$, застосованого всіма учасниками гри. Функції f_k також називають **функціями виграшу**, або **платіжними функціями**. У тому випадку, якщо для будь-яких S

$$\sum_{k=1}^{K} f_k(s) = 0,$$

гра називається **грою з нульовою сумою**. Гру із двома учасниками й нульовою сумою називають **антагоністичною**. Антагоністичні ігри, тобто ігри, у яких виграш одного учасника дорівнює програшу іншого, у зв’язку з відносно простою постановкою задачі є найбільш вивченим розділом теорії ігор. Однак зміст теорії ігор, безумовно, не вичерпується ними. У класифікації ігрових моделей виділяють ігри з кінцевими й нескінченними наборами стратегій у гравців, виділяють ігри за можливими кількостями ходів учасників. Також ігри поділяють на некооперативні й кооперативні, тобто ті, в яких функції виграшу учасників залежать від утворених ними коаліцій. Крім цього ігри можна розрізняти за обсягом інформації, наявної в гравців щодо минулих ходів. У цьому зв'язку їх поділяють на ігри з повною й неповною інформацією.

Розглянемо докладніше антагоністичні ігри і їх основні властивості. Зручним способом завдання гри двох учасників з нульовою сумою є **платіжна матриця**. Звідси, до речі, випливає ще одна їх назва - **матричні ігри**. Кожний елемент платіжної матриці a_{ij} містить числове значення виграшу гравця *I* (програшу гравця *II*), якщо перший застосовує стратегію *і*, а другий - стратегію *j*. Терміни «виграш» і «програш» треба розуміти в широкому змісті, тому що вони можуть приймати від’ємні значення й означати протилежне. Нетривіальність задачі, насамперед, полягає в тому, що кожний із гравців робить свій вибір, не знаючи про вибір іншого, що істотно ускладнює процес оптимізації обираної стратегії.

Класичним прикладом антагоністичної гри є гра із двома учасниками, що загадують незалежно один від одного числа. Передбачається, що якщо їх сума опиняється парною, то виграш, рівний 1, дістається першому гравцеві, а якщо непарною, то другому. Поклавши, що для обох гравців загадування непарного числа є першою стратегією, а парного - другою, можемо записати платіжну матрицю даної гри:

$$\begin{array}{c|cc} & \textbf{н/п} & \textbf{п} \\ \hline \textbf{н/п} & 1 & -1 \\ \textbf{п} & -1 & 1. \end{array} \qquad (10.1)$$

Рядки матриці (10.1) відповідають стратегіям гравця I, стовпці - стратегіям гравця II, а її елементи - результатам першого гравця. Також з визначення гри випливає, що елементи цієї матриці, узяті зі зворотним знаком, відповідають виграшам другого гравця.

Складніша й змістовніша платіжна матриця може бути отримана, якщо трохи модифікувати запропоновану гру. Припустимо, що обоє учасники мають право загадувати числа від 1 до 4, що є їх відповідними стратегіями. У випадку якщо результат додавання задуманих чисел буде парним, то другий гравець виплачує першому суму, що вийшла, а якщо непарним, то перший - другому. Запишемо платіжну матрицю для такої гри:

$$A=\begin{array}{cccc} \left|\begin{array}{cccc} 2 & -3 & 4 & -5 \\ -3 & 4 & -5 & 6 \\ 4 & -5 & 6 & -7 \\ -5 & 6 & -7 & 8 \end{array}\right| \begin{array}{c} 1 \\ 2 \\ 3 \\ 4 \end{array} \\ \begin{array}{cccc} 1 & 2 & 3 & 4 \end{array} \end{array} \qquad (10.2)$$

Як ми вже відзначали, найважливішим у теорії ігор є питання про оптимальність розв’язку (вибору стратегії) для кожного із гравців. Проаналізуємо з цього погляду певну матричну гру, для якої задана платіжна матриця $A = \left\| a_{ij} \right\|_{m \times n}$. При виборі гравцем *I* стратегії *i* його гарантований дохід незалежно від дій гравця *II* складе $\min\limits_j a_{ij}$. Оскільки він може вибирати *i* самостійно, то доцільно цей вибір зробити таким, щоб він при будь-якій стратегії супротивника максимізував величину гарантованого доходу, тобто забезпечував одержання $\max\limits_i \left(\min\limits_j a_{ij} \right)$. Такий принцип вибору стратегії одержав назву **принцип максиміна**, а розмір гарантованого виграшу – **нижньої ціни гри**. З іншого боку, аналогічні міркування можна провести із приводу дій другого гравця. Його найбільший програш при виборі стратегії *j* складе $\max\limits_i a_{ij}$, і, отже, йому треба вибирати стратегію так, щоб мінімізувати величину програшу при будь-яких діях суперника, тобто забезпечити $\min\limits_j \left(\max\limits_i a_{ij} \right)$. у цьому полягає суть **принципу мінімакса**, розмір програшу називається **верхньою ціною гри**.

Можна довести справедливість наступного співвідношення:

$$\max_i \min_j a_{ij} \le \min_j \max_i a_{ij} . \tag{10.3}$$

Проте очевидний інтерес представляє ситуація, при якій значення виграшу (платежу), одержуваного гравцем *I* при виборі ним максимінної стратегії, дорівнює платежу (програшу) *П*-го гравця при мінімаксній стратегії

$$\max_i \min_j a_{ij} = \min_j \max_i a_{ij} . \tag{10.4}$$

У цьому випадку говорять, що гра має **сідлову точку**. Збіг значень гарантованих виграшів гравців при максимінній і мінімаксній стратегії означає можливість досягнення в грі певного оптимального (стабільного, рівноважного) стану, від якого невигідно відхилятися жодному з учасників. Поняття «оптимальність» тут означає, що жоден розумний (обережний) гравець не прагне змінити свою стратегію, тому що його супротивник, у принципі, зможе вибрати таку стратегію, що дасть гірший для першого результат. Стратегії i^* і j^*, що утворюють сідлову точку, називають **оптимальними**, а значення $v = a_{i^* j^*}$ називають **ціною гри**. Трійку (i^*,j^*,v) вважають **розв’язком матричної гри із сідловою точкою**.

Неважко помітити, що не усяка гра має сідлову точку. Зокрема, як гра (10.1), так і гра (10.2) сідлової точки не мають. Прикладом гри, що має сідлову точку, є гра із платіжною матрицею (10.5):

$$A = \begin{vmatrix} 7 & 10 & 4 & 1 \\ 6 & 8 & 5^* & 17 \\ 8 & -3 & 2 & 11 \end{vmatrix} \tag{10.5}$$

У цій матриці мінімальні (гарантовані) виграші першого гравця за рядками дорівнюють 1, 5 і (-3). Отже, його максимінному вибору відповідатиме стратегія 2, що гарантує виграш 5. Для другого гравця максимальні програші за стовпцями матриці складуть 8, 10, 5, 17, тому має смисл зупинитися на стратегії 3, при якої він програє тільки 5. Отже, друга стратегія першого гравця і третя стратегія другого утворюють сідлову точку зі значенням 5, тобто гра з матрицею (10.5) має розв’язання (2; 3; 5).

10.3 Змішані стратегії. Основна теорема теорії ігор

Подальший розвиток теорії матричних ігор ґрунтується на дослідженні гри як певного повторюваного процесу. Дійсно, навряд чи можна дати змістовні рекомендації з такого питання, як треба поводитися учасникам однократно проведеної гри, що не має сідлової точки. У випадку її багаторазових повторів природною і плідною уявляється ідея рандомізації вибору стратегій гравцями, тобто внесення до процесу вибору елемента випадковості. Дійсно, систематичне відхилення, наприклад, гравця *I* від максимінної стратегії з метою збільшення виграшу може бути зафіксовано другим гравцем і покарано. У той самий час абсолютно хаотичний вибір стратегій не принесе в середньому найкращого результату.

Змішаною стратегією гравця *I* у грі з матрицею $A = \left\| a_{ij} \right\|_{m \times n}$ називають впорядкований набір дійсних чисел x_i, $i = \overline{1, m}$, що задовольняють умовам

$$x_i \ge 0, \quad i = \overline{1, m}; \quad \sum_{i=1}^{m} x_i = 1. \tag{10.6}$$

Числа x_i інтерпретують як імовірності застосування гравцем *I* стратегій 1,2, …, *m*, які, на відміну від змішаних, називають **чистими стратегіями**.

Аналогічно вводять поняття змішаних стратегій гравця *II*, які визначають як набір чисел y_j, $j = \overline{1, n}$, що задовольняють умовам

$$y_j \ge 0, \quad j = \overline{1, n}; \quad \sum_{j=1}^{n} y_j = 1. \tag{10.7}$$

Тоді, якщо гравець *I* застосовує змішану стратегію $x=(x_1, x_2, ..., x_m)$, а гравець *II* змішану стратегію $y = (y_1, y_2, \dots, y_n)$, математичне сподівання виграшу гравця *I* (програшу гравця *II*) визначається співвідношенням

$$F(x, y) = \sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij} x_i y_j . \tag{10.8}$$

Надалі через *X* позначимо множину припустимих змішаних стратегій гравця *I*, зумовлену умовою (10.6), а через *Y* - зумовлену умовою (10.7) множину припустимих змішаних стратегій гравця *II*.

До пошуку розв’язку гри в змішаних стратегіях можуть бути застосовані критерії максиміна-мінімакса. Відповідно до них гравець *I* вибиратиме свою змішану стратегію $x=(x_1, x_2, ..., x_m)$ так, щоб максимізувати найменший середній виграш:

$$\max_{x\in X}\left[\min_{y\in Y}\sum_{i=1}^{m}\sum_{j=1}^{n}a_{ij}x_i y_j\right], \tag{10.9}$$

який дорівнює

$$\max_{x\in X}\left[\min\left\{\sum_{i=1}^{m}a_{i1}x_i, \sum_{i=1}^{m}a_{i2}x_i, ..., \sum_{i=1}^{m}a_{in}x_i\right\}\right], \tag{10.10}$$

а гравець *II* - свою змішану стратегію так, щоб мінімізувати найбільший середній програш:

$$\min_{y\in Y}\left[\max_{x\in X}\sum_{i=1}^{m}\sum_{j=1}^{n}a_{ij}x_i y_j\right], \tag{10.11}$$

який дорівнює

$$\min_{y\in Y}\left[\max\left\{\sum_{j=1}^{n}a_{1j}y_j, \sum_{j=1}^{n}a_{2j}y_j, ..., \sum_{j=1}^{n}a_{mj}y_j\right\}\right]. \tag{10.12}$$

За аналогією з (10.3) для будь-яких $x\in X$ і $y\in Y$ справедливою є нерівність

$$\max_{x\in X}\left[\min_{y\in Y}\sum_{i=1}^{m}\sum_{j=1}^{n}a_{ij}x_i y_j\right]\le\min_{y\in Y}\left[\max_{x\in X}\sum_{i=1}^{m}\sum_{j=1}^{n}a_{ij}x_i y_j\right]. \tag{10.13}$$

Стратегії $x^*\in X$ і $y^*\in Y$ називають **оптимальними змішаними стратегіями**, якщо для будь-яких $x\in X$ і $y\in Y$ справедлива рівність

$$F(x^*, y^*)=\max_{x\in X}\left[\min_{y\in Y}\sum_{i=1}^{m}\sum_{j=1}^{n}a_{ij}x_i y_j\right]=\min_{y\in Y}\left[\max_{x\in X}\sum_{i=1}^{m}\sum_{j=1}^{n}a_{ij}x_i y_j\right]. \tag{10.14}$$

$v = F(x^*, y^*)$ називають ціною гри, і якщо x^* і y^* існують, говорять, що **гра має розв’язання в змішаних стратегіях (*x*, в*, v*)**.

Справедливою є фундаментальна теорема Дж. Неймана, (приведемо без доказу).

Теорема 10.1 (основна теорема матричних ігор). Будь-яка матрична гра має розв’язання в змішаних стратегіях.

Значення і нетривіальність теореми (10.1) зумовлені насамперед тим, що, як було показано раніше, у загальному випадку матричні ігри в чистих стратегіях розв’язання не мають.

10.4 Зведення гри двох гравців до задачі лінійного програмування

Задача, розв'язувана першим гравцем, (10.10) була сформульована як максимізація найменшої із сум

$$\sum_{i=1}^{m}a_{ij}x_i ,$$

але якщо визначити певне x_{m+1}, для якого виконується

$$x_{m+1} \le \sum_{i=1}^{m} a_{ij} x_i \ , \tag{10.15}$$

її можна звести до задачі лінійного програмування:

знайти $$F(x) > \max \tag{10.16}$$

при обмеженнях

$$\sum_{i=1}^{m} a_{ij} x_i \ge x_{m+1}, \quad j = \overline{1,n}; \quad \sum_{i=1}^{m} x_i = 1; \quad x_i \ge 0; \quad i = \overline{1,m}. \tag{10.17}$$

Провівши аналогічні міркування, приходимо до того, що задача мінімізації найбільшого сподіваного програшу, розв'язувана гравцем *II* (10.12), зводиться до задачі лінійного програмування

$$F(x) > \min, \tag{10.18}$$

$$\sum_{j=1}^{n} a_{ij} y_j \le y_{n+1}, \quad i = \overline{1,m}; \quad \sum_{j=1}^{n} y_j = 1; \quad y_j \ge 0; \quad j = \overline{1,n}. \tag{10.19}$$

Отже, ми одержуємо можливість застосовувати всі можливості апарата лінійного програмування для пошуку оптимальних стратегій обох гравців.

Досить легко перевірити, що задачі (10.16)-(10.17) і (10.18)-(10.19) утворюють двоїсту пару. Тут у певному змісті ми повернулися до взаємозв'язку між наявністю розв’язку в певної оптимізаційної задачі й існуванням сідлової точки у відповідної функції Лагранжа. У цьому випадку аналогічний зв'язок простежується між сідловою точкою гри й розв’язком пари задач оптимізації.

Графічний метод вирішення гри. Слід зазначити, що застосування для розв’язання задач (10.16)-(10.17), (10.18)-(10.19) стандартних алгоритмів лінійного програмування далеко не завжди є раціональним. Крім цього існують інші методи, які ґрунтуються на використанні специфіки цих задач. Зупинимося на дуже простому класичному способі пошуку оптимальних змішаних стратегій у матричних іграх, де один з учасників має тільки дві стратегії (це так звані $2\times n$ і $m\times 2$ ігри).

Для визначеності покладемо, що гравець *I* має можливість вибирати між двома стратегіями з імовірностями x_1 і x_2=l-x_1, тоді його очікувані виграші, що відповідають чистим стратегіям гравця *II*, приймуть вигляд

$$a_{11}x_1 + a_{21}(1-x_1),\ a_{12}x_1 + a_{23}(1-x_1),...,a_{1n}x_1 + a_{2n}(1-x_1)\,,$$

або

$$(a_{11}-a_{21})x_1 + a_{21},\ (a_{12}-a_{22})x_1 + a_{22},...,(a_{1n}-a_{2n})x_1 + a_{2n}\,,$$

тобто очікувані виграші можуть бути представлені у вигляді графіків лінійних функцій, що залежать від змінної x_1 (рис. 10.1), де передбачається, що гравець *II* має три стратегії).

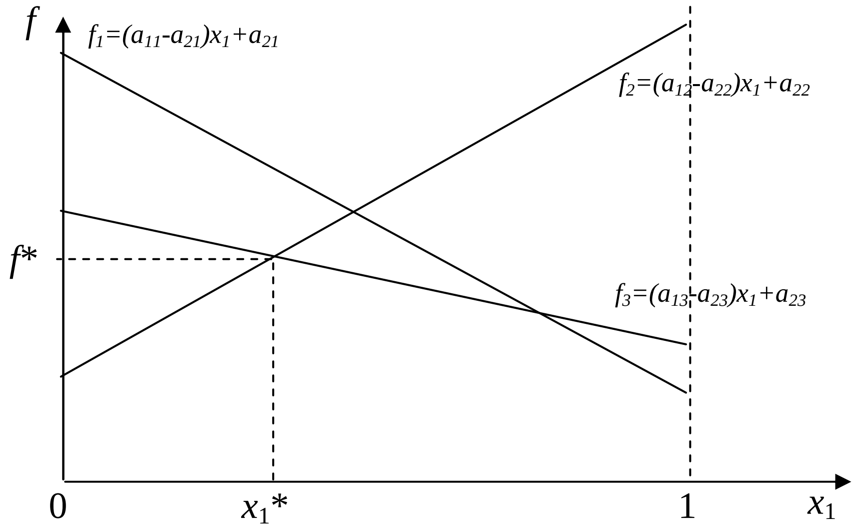


Рисунок 10.1 - Графічне вирішення гри

Лінії, зображені на рисунку 10.1, задають залежності середнього виграшу гравця *I* від значення ймовірності x_1, з якою він вибирає свою першу стратегію, для випадків, коли його супротивник вибирає першу, другу або третю чисту стратегію. Тоді значенням мінімального гарантованого доходу першого гравця відповідає нижня огинаюча всіх трьох прямих. Відповідно до принципу максиміна, оптимальному вибору гравця *I* відповідатиме найвища точка, що лежить на даній огинаючій, позначена на рисунку як (x_1*, f*). Знаючи її, можна визначити оптимальну змішану стратегію першого гравця $x^*=(x_1^*,\ 1\text{-}x_2^*)$ і ціну гри, що дорівнює f*.

Виходячи з відношення подвійності, яким пов'язані задачі обох гравців, за оптимальною стратегією першого учасника x* однозначно визначається оптимальна стратегія його супротивника y*. Оскільки y є результатом розв’язання задачі лінійного програмування, то він має всі властивості припустимого базисного плану, тобто у випадку $2\times n$ гри має не більше за дві ненульових компоненти і не менше за (n-2) нульових. Номера ненульових елементів y* визначаються номерами ліній, перетинання яких визначило оптимальну стратегію першого гравця. Дійсно, гравець *II* знає оптимальну стратегію суперника, і застосування ним стратегій, що відповідають прямим, які проходять вище точки (x_1*, f*), тільки збільшить його програш. У розглянутому прикладі це лінії f_2 і f_3, і, отже, у своїй оптимальній стратегії другий гравець повинен з ненульовими ймовірностями застосовувати другу і третю чисті стратегії ($y_2>0$, $y_3>0$). На основі цього, а також з огляду на умову нормування

$$\sum_{j=1}^{n} y_j = 1,$$

виразимо: $y_3=1-y_2$, тоді оптимальне значення y^*_2 можна знайти з умови

$$a_{11}\times 0 + a_{21}y_2^* + a_{31}(1-y_2^*) = a_{12}\times 0 + a_{22}y_2^* + a_{32}(1-y_2^*)\,,$$

або

$$(a_{21}-a_{31})y_2^* + a_{31} = (a_{22}-a_{32})y_{21}^* + a_{32}\,.$$

В результаті одержуємо оптимальну стратегію гравця *II* $y^*=(0, y^*_2, y^*_3)$.

Очевидно, що пошук розв'язання в грі $m\times2$ здійснюється аналогічно: будують графіки очікуваного програшу гравця *II*, знаходять їх верхню огинаючу та ін.

Безумовно, графічний спосіб в силу обмеженості кола задач, до яких він може бути застосований, має скоріше теоретичне, ніж практичне значення. Однак він добре ілюструє змістовну сторону процесу пошуку розв'язку в грі.

Контрольні запитання

1. Коротко сформулюйте предмет теорії ігор як наукової дисципліни.
2. Який зміст вкладається в поняття «гра»?
3. Для опису яких економічних ситуацій може бути застосований апарат теорії ігор?
4. Яка гра називається антагоністичною?
5. Чим однозначно визначаються матричні ігри?
6. У чому полягають принципи максиміна й мінімакса?
7. За яких умов можна говорити про те, що гра має сідлову точку?
8. Наведіть приклади ігор, які мають сідлову точку й у яких вона відсутня.
9. Які підходи існують до визначення оптимальних стратегій?
10. Що називають «ціною гри»?
11. Дайте визначення поняттю «змішана стратегія».
12. Сформулюйте основну теорему матричних ігор.

ТЕМА 11
КРИТЕРІЇ ОПТИМАЛЬНОСТІ В УМОВАХ ПОВНОЇ НЕВИЗНАЧЕНОСТІ

Невизначеність, пов'язану з відсутністю інформації про імовірності станів середовища (природи), називають«безнадійною» або «дурною». У таких випадках для визначення найкращих рішень використовують ряд критеріїв:

- критерій гарантованого результату (максимінний критерій Вальда). Це песимістичний за своєю сутністю критерій, тому що бере до уваги тільки найгірший із всіх можливих результатів кожної альтернативи. Такий підхід встановлює гарантований мінімум;
- критерій оптимізму (критерій максимакса) відповідає оптимістичній стратегії; тут не приймається до уваги ніякий можливий результат, крім найкращого;
- критерій песимізму характеризується вибором гіршої альтернативи з гіршим із всіх гірших значень окупності;
- критерій мінімаксного ризику Севіджа можна розглядати як критерій найменшої шкоди, що визначає гірші можливі наслідки для кожної альтернативи й вибирає альтернативу з кращим із поганих значень;
- критерій узагальненого максиміна (песимізму - оптимізму) Гурвіца дозволяє враховувати стан між крайнім песимізмом і невтримним оптимізмом.

У певних обставинах кожний з цих методів має свої переваги й недоліки, які можуть допомогти у прийнятті рішення. При порівняльному аналізі критеріїв ефективності недоцільно зупинятися на виборі єдиного критерію, тому що в ряді випадків це може призвести до невиправданих рішень, які ведуть до значних втрат економічного, соціального та іншого змісту. Тому в зазначених ситуаціях є необхідність застосування кількох критеріїв у сукупності. Наприклад, поряд з критерієм гарантованого результату може бути використаний критерій Севіджа, критерій оптимальної поведінки може доповнюватися застосуванням песимістичного критерію та ін.

Застосування різних критеріїв ефективності для різних задач вибору оптимальних рішень в умовах невизначеності показує, що підхід, який ґрунтується на комплексному застосуванні зазначених критеріїв, може стати визначальним.

11.1 Критерій гарантованого результату

Критерій гарантованого результату також називають максимінним критерієм Вальда. Сутність цього критерію полягає в наступному. Є множина стратегій (варіантів, альтернатив) рішення проблеми $P = \{P_i\}$, $i = \overline{1, m}$. Зазначені стратегії вважають керованими факторами.

Поряд з факторами керованими діють фактори, які не піддаються управлінню. Позначимо їх $П=\{П_j\}$, $j = \overline{1, n}$. Фактори $П_j$ можуть бути рівнем попиту на товари, пропоновані фірмою, ринковими цінами, діями конкурентів та ін.

Для оцінки ефективності прийнятих рішень вводять показник ефективності E і вважають, що функція $E(P, П)$ є відомою. Оскільки фактори P і $П$ дискретні, ефективність E є множиною дискретних чисел. Отже, кожній точці контрольованих і неконтрольованих факторів $(P_i, П_j)$, ставиться у відповідність значення ефективності $E(P_i, П_j)$. Матриця $E = \|e_{ij}\|$ представлена у вигляді таблиці 11.1.

Для кожного контрольованого фактору P_i знаходять $\min\limits_{1 \le j \le n}\{e(P_i, П_j)\}$, у результаті чого отримують набір значень показника ефективності. Потім вибирають керований фактор P_i, при якому забезпечується максимальне значення $E(P, П)$.

Таблиця 11.1 - Матриця ефективності

$П_j$ / P_i	$П_1$	$П_2$	...	$П_n$	**min** e_{ij}
P_1	e_{11}	e_{12}	...	e_{1n}	$e(P_1, П)$min
P_2	e_{21}	e_{22}	...	e_{2n}	$e(P_2, П)$min
…	…	…	…	…	…
P_m	e_{m1}	e_{m2}	...	e_{mn}	$e(P_m, П)$min

Отже, критерій гарантованого результату (максимінний критерій Вальда) записують у вигляді

$$E_{Г} = \max_{i}\min_{j} E(P, П). \qquad (11.1)$$

Цей критерій забезпечує максимізацію мінімального виграшу або, що те саме, мінімізацію максимальних втрат, які можуть бути при реалізації однієї із стратегій. Критерій простий і чіткий, але консервативний у тому розумінні, що орієнтує на занадто обережну лінію поведінки. Величину, що відповідає максимальному критерію, називають нижньою ціною гри, під якою треба розуміти максимальний виграш, що гарантований у грі з певним супротивником вибором однієї із своїх стратегій при мінімальних результатах. Це перестрахувальна позиція крайнього песимізму, розрахована на гірший випадок. Така стратегія прийнятна, наприклад, коли гравець не настільки зацікавлений у великій удачі, але бажає себе застрахувати від несподіваних програшів. Вибір такої стратегії визначається ставленням гравця до ризику.

У ряді економічних задач як критерій ефективності прийнятих рішень виступає показник мінімуму витрат. У цьому випадку принцип гарантованих витрат формулюється у вигляді

$$З_{Г} = \min_{i}\max_{j} З_{ij}. \qquad (11.2)$$

Витратами $З = \| З_{ij} \|$ можуть бути капітальні вкладення, валові витрати виробництва, приведені річні витрати й інші показники.

Розглянемо приклад. Провадиться порівняння різних інвестиційних проектів $Пр_1$, $Пр_2$,..., $Пр_m$. Для реалізації кожного з проектів необхідна певна величина капітальних вкладень $K=\{K_i\}$, величини K_i є керованими (контрольованими) факторами.

Кожному проекту відповідає певне значення собівартості продукції, що передбачається випускати при реалізації проекту. Сукупність значень собівартості продукції представлена у вигляді $C=\{C_j\}$.

Величини C_j на початкових етапах виконання проекту точно визначити неможливо, тому їх вважають неконтрольованими факторами. Кожній парі (K_i, C_j) відповідає певне значення приведених річних витрат, що визначають за формулою

$$З_{ij} = E_н * K_i + C_j,$$

де $E_н$ - нормативний коефіцієнт ефективності капітальних вкладень.

Маючи у своєму розпорядженні набори $\{K_i\}$ і $\{C_j\}$, складемо матрицю приведених витрат $З = \| З(K_i, C_j) \| = \| З_{ij} \|$ (табл. 11.2).

Таблиця 11.2 - Залежність приведених витрат від K і $З$

C / **K**	$\boldsymbol{C_1}$	$\boldsymbol{C_2}$	$\boldsymbol{C_3}$	$\boldsymbol{C_4}$	**max** $\boldsymbol{C_{ij}}$
K_1	100	130	75	90	130
K_2	80	200	140	160	200
K_3	60	180	200	100	200
K_4	130	90	150	150	150

Критерій гарантованих витрат реалізується як

$$З_Г = \min\{130,200,200;150\}=130.$$

Як найефективніша виступає перша стратегія, який відповідають капітальні вкладення K_1.

11.2 Критерій оптимізму

При використанні цього критерію, називаного також критерієм максимакса, припускається, що умови функціонування аналізованих систем будуть найбільш сприятливими. Внаслідок цього оптимальним рішенням є стратегія, що призводить до одержання найбільшого значення критерію оптимальності в платіжній матриці. Цей критерій доцільно застосовувати в тих випадках, коли є принципова можливість вплинути на функції протилежної сторони.

При аналізі матриці ефекту $E(P, П)$ того або іншого виду вибір керованих факторів здійснюється так, щоб забезпечити максимальний ефект. Критерій оптимізму записують у вигляді

$$E_0 = \max_i \max_j E(P, П). \qquad (11.3)$$

Якщо розглядають матрицю витрат, керовані фактори вибирають так, щоб мінімізувати зазначені витрати. Тоді розглянутий критерій формулюється в такий спосіб:

$$З_0 = \min_i \min_j З(P, П). \qquad (11.4)$$

Розглянемо приклад. Розглядають матрицю приведених річних витрат, що відповідає таблиці 11.2. Необхідно визначити найефективнішу стратегію, використовуючи критерій оптимізму. Стосовно до розглянутої ситуації принцип оптимізму можна представити у вигляді

$$З_0=\min\{75,80,60,90\}= 60.$$

Отже, найефективнішою є стратегія, що відповідає K_3. Порівнюючи два рішення (критерій Вальда та критерій оптимізму), бачимо, що вони не збігаються. Така ситуація буде характерною для більшості аналізованих реальних задач через принципові відмінності критеріїв.

11.3 Критерій песимізму

На відміну від критерію оптимізму, при використанні принципу песимізму припускають, що керовані фактори можуть бути використані несприятливим способом:

$$E_n = \min_i \min_j E(P, \Pi). \tag{11.5}$$

У реальних ситуаціях у ряді задач може виявитися неможливим контроль за неконтрольованими факторами, що належать множині *P*. Особливо це стосується задач, що пов'язані з необхідністю урахування фактору часу. До цих задач можна віднести соціально-економічне прогнозування, довгострокове планування, проектування складних об'єктів та ін. Наприклад, витрати виробництва є контрольованими факторами на коротких часових інтервалах. Однак, при аналізі тривалих процесів, які складають кілька років, певні елементи зазначених витрат стають неконтрольованими. До них можна віднести вартість електроенергії, вартість матеріалів і покупних виробів та ін.

Для аналізу матриці витрат критерій песимізму запишеться як

$$З_n = \max_i \max_j З(P, \Pi).$$

Розглянемо приклад. Маючи у своєму розпорядженні матрицю приведених річних витрат, представлену у вигляді таблиці 11.2, необхідно вибрати ефективну стратегію за допомогою принципу песимізму.

У розглянутій стратегії $З_{\text{п}}$ = max{130,200,200,150}= 200.

Витрати $З_{\text{п}}$ = 200 можна забезпечити при використанні другої і третьої стратегій.

11.4 Критерій мінімаксного ризику Севіджа

При використанні розглянутих критеріїв можливі ситуації, коли неконтрольовані фактори діятимуть сприятливіше порівняно з найгіршим станом. Наприклад, погодні умови можуть виявитися сприятливішими порівняно з прогнозованими. Число конкурентів на ринках може виявитися істотно меншим за прогнозоване.

У подібних ситуаціях корисний результат може значно відрізнятися від того, який був отриманий за допомогою критерію гарантованого результату або критерію песимізму. Тому виникає необхідність оцінки можливих відхилень результативної ознаки від оптимальних значень. Тут знаходить застосування критерій Севіджа. Вибір стратегії аналогічний вибору стратегії за принципом Вальда з тією відмінністю, що гравець керується не матрицею виграшів *E*, а матрицею ризиків *R*, яку будують за формулою

$$r_{ij} = \left(\max_{1 \le i \le m} e_{ij}\right) - e_{ij}. \tag{11.6}$$

Критерій Севіджа формулюють в такий спосіб:

$$E_{rc} = \min_i \max_j R(P, \Pi). \tag{11.7}$$

Сутність цього критерію полягає в прагненні уникнути великого ризику при виборі рішення. Отже, критерій Севіджа мінімізує можливі втрати. Основним вихідним припущенням цього критерію є припущення про те, що на вибір варіантів впливають дії розумних супротивників, інтереси яких прямо протилежні інтересам ОПР.

11.5 Критерій узагальненого максиміна Гурвіца

Критерій Гурвіца дозволяє враховувати комбінації найгірших станів. Цей критерій при виборі рішення рекомендує керуватися певним середнім результатом, що характеризує стан між крайнім песимізмом і невтримним оптимізмом.

Відповідно до цього компромісного критерію для кожного рішення визначають лінійну комбінацію мінімального й максимального виграшів.

$$E_i = k\{\min_j e_{ij} + (1-k)\max_j e_{ij}\},$$

і перевагу віддають варіанту рішення, для якого виявиться максимальним показник E_i тобто

$$E_{i\Gamma} = \max_i\{k\min_j e_{ij} + (1-k)\max_j e_{ij}\}, \qquad (11.8)$$

де k - коефіцієнт, що розглядається як показник оптимізму ($0 \le k \le 1$).

При $k=0$ критерій Гурвіца збігається з максимальним критерієм, тобто орієнтація на граничний ризик, тому що більший виграш сполучений, як правило, з більшим ризиком. При $k=1$ - орієнтація на обережну поведінку. Значення k між 0 і 1 є проміжними між ризиком і обережністю та вибираються залежно від конкретної обстановки і схильності до ризику ОПР.

Щодо матриці ризиків R критерій Гурвіца має вигляд:

$$E_{r\Gamma} = \min_i\left\{k\max_j r_{ij} + (1-k)\min_j r_{ij}\right\}. \qquad (11.9)$$

Зведемо всі критерії оптимальності до таблиці 11.3.

Таблиця 11.3 - Критерії оптимальності

Показник	**Формула**	**Назва**
Найбільша обережність	$E_г = \max \min e_{ij}$	Критерій гарантованого результату (Вальда)
Найменша обережність	$E_г = \max \max e_{ij}$	Критерій оптимізму
Крайня обережність	$E_п = \min \min e_{ij}$	Критерій песимізму
Мінімальний ризик	$E_{гс} = \min \max r_{ij}$	Критерій Севіджа

Продовження таблиці 11.3

Показник	Формула	Назва
Компроміс у рішенні	$E_{iГ} = \max\limits_i \{k \min\limits_j e_{ij} + (1-k)\max\limits_j e_{ij}\}$ або $E_{rГ} = \min\limits_i \left\{ k\max\limits_j r_{ij} + (1-k)\min\limits_j r_{ij} \right\}$	Критерій Гурвіца

11.6 Порівняльна оцінка варіантів рішень залежно від критеріїв ефективності

За числом критеріїв оцінки альтернатив виділяють одно- і багатокритеріальні задачі прийняття рішення. Принципова різниця між ними полягає в тому, що в умовах багатокритеріальності виникає проблема порівняння, сукупного урахування вимог різних критеріїв, що на відміну від задачі впорядкування альтернатив за одним критерієм не може бути вирішеною формальним шляхом і вимагає звертання до ОПР. Необхідно мати на увазі, що будь-якого формального математичного методу «подолання» багатокритеріальності не може існувати в принципі. Всі без винятку методи рішення багатокритеріальних задач є різними способами організації взаємодії з ОПР.

Наявність кількох критеріїв вибору ефективних альтернатив вносить додаткову невизначеність при прийнятті найкращих рішень. Отже, має місце невизначеність двох видів:

1) невизначеність, зумовлена відсутністю або нестачею інформації про аналізовані процеси;

2) невизначеність, причиною якої є наявність кількох принципів оптимальності.

Нехай при виборі ефективних рішень при наявності некерованих факторів використовують множину критеріїв оптимальності $G = \{G_i\}$, $i = \overline{1,m}$. Критерії G_i є функцією керованих факторів $P = \{P_i\}$, і некерованих факторів $П = \{П_i\}$. Маючи у своєму розпорядженні множину критеріїв $G = \{G(P, П)\}$, $i = \overline{1,m}$ необхідно вибрати ефективне рішення з урахуванням зазначеної сукупності рішень.

У випадку відсутності інформації про імовірності станів середовища теорія не дає однозначних і математично строгих рекомендацій з вибору критеріїв прийняття рішень. Це пояснюється у великій мірі не слабкістю теорії, а невизначеністю самої ситуації. Єдиний розумний вихід у подібних випадках - спробувати отримати додаткову інформацію, наприклад, шляхом проведення досліджень або експериментів. Під час відсутності додаткової інформації прийняті рішення є теоретично недостатньо обґрунтованими й значною мірою суб'єктивними. Хоча застосування математичних методів в іграх з природою не дає абсолютно достовірного результату і останній певною мірою є суб'єктивним, проте воно створює певне впорядкування наявних у розпорядженні ОПР даних: задається множина станів природи, альтернативні рішення, виграші й втрати при різних сполученнях стану «середовище-рішення». Таке впорядкування уявлень про проблему сприяє підвищенню якості прийнятих рішень.

Оптимальність за Парето. Як було відзначено, аналіз рішень при багатьох критеріях у значній мірі зводиться до організації в тій або іншій формі взаємодії з ОПР, що одна тільки й може розв'язати проблему порівняння різних критеріїв. Проте, існує досить обмежена область, у якій застосування сугубо формального аналізу без звертання до ОПР опиняється досить корисним. Мова йде про виділення так званої множини ефективних, або оптимальних за Парето, альтернатив.

Очевидно, що альтернатива, яка не є ефективною ні при яких умовах, не може розглядатися як рішення задачі. Звідси випливає найважливіший **критерій раціональності** процесу розробки рішення: обираний варіант повинен бути ефективним.

Ефективною вважають таку альтернативу, для якої не існує іншої припустимої, що не уступає їй за всіма критеріями і хоча б за одним критерієм переважаючої її.

Головне у відшуканні ефективного рішення полягає в тому, що після того, як сформульовані критерії, задача відшукання множини ефективних рішень на заданій множині альтернатив є формальною задачею (що не потребує для рішення звертання до ОПР). У багатьох випадках для визначення множини ефективних альтернатив використовують інтегральний критерій оптимальності, який представляє собою суму окремих, частинних критеріїв із змінними вагами. При цьому перебирають з заданим кроком всі можливі комбінації вагових коефіцієнтів на відрізку від 0 до 1. Можна для визначення вагових коефіцієнтів (коефіцієнтів значущості) звернутися до ОПР або скористуватися для їх визначення методом експертних оцінок. Потім будують так звану згортку критеріїв, тобто як інтегральний показник якості альтернативи приймають суму окремих критеріїв з ваговими коефіцієнтами. Таку методику використовують дуже часто. До її переваг, крім простоти, треба віднести те, що одержувана при такому підході альтернатива свідомо буде ефективною. Однак застосування цієї схеми засноване на додаткових припущеннях, які не завжди виправдані. З математичної точки зору така сума частинних критеріїв з ваговими коефіцієнтами є адитивною функцією цінності. Для того, щоб така логічна конструкція правильно відображала систему переваг ОПР, необхідно (про це доведені відповідні теореми), щоб використовувані для оцінки альтернатив критерії мали властивість взаємної незалежності.

Розглянемо приклад. Нехай є матриця платоспроможного попиту E і відповідна їй матриця ризиків R

$$E = \begin{pmatrix} 49300 & 197200 & 197200 & 197200 \\ -60 & 148900 & 297800 & 297800 \\ -1140 & 98400 & 196800 & 393600 \end{pmatrix}, \qquad R = \begin{pmatrix} 0 & 0 & 100600 & 196400 \\ 49360 & 48300 & 0 & 95800 \\ 50440 & 988800 & 101000 & 0 \end{pmatrix}.$$

Припустимо, що відомі імовірності P_j того, що ситуація розвивається за варіантом j. Таке положення називають частковою невизначеністю. Тоді рішення можна приймати, зокрема, за правилом максимізації середнього очікуваного доходу.

Прибуток, одержуваний компанією при реалізації i-го рішення, є випадковою величиною E_i з рядом розподілу, що наведений у таблиці 11.4.

Таблиця 11.4 - Ряд розподілу випадкової величини E_i

E_i	e_{i1}	e_{i2}	…	e_{in}
P_i	P_1	P_2	…	P_n

Середній прибуток є математичним сподіванням $M[E_i]$ або $\overline{E}_i$. Отже, правило рекомендує прийняти рішення, що приносить максимальний середній прибуток. Нехай імовірності P_j дорівнюють: 1/6, 1/4, 1/4, 1/3. Тоді

$\overline{E}_1$ = 49300*1/6 + 197200*1/4 + 197200*1/4 + 197200*1/3 = 172500,

$\overline{E}_2$ = -60*1/6+148900*1/4+297800*1/4 + 297800*1/3 = 210931,

$\overline{E}_3$ = -1140*1/6 + 98400*1/4 + 196800*1/4 + 393600*1/3 = 204810.

Максимальний середній прибуток дорівнює $\overline{E}_2$ = 210931 і відповідає стратегії компанії P_2.

Далі розглянемо вибір рішення за правилом мінімізації середнього ризику. Ризик компанії при реалізації i-го рішення є випадковою величиною R, з рядом розподілу, що наведений у таблиці 11.5.

Таблиця 11.5 - Ряд розподілу випадкової величини R

R_i	r_{i1}	r_{i2}	…	r_{in}
P_i	P_1	P_2	…	P_n

Середній ризик є математичним сподіванням $M[R_i]$ або $\overline{R}_i$. Правило рекомендує прийняти рішення, що забезпечує мінімальний середній ризик.

Обчислимо середні ризики при зазначених вище імовірностях для матриці ризиків R. Одержимо:

$\overline{R}_1$ = 0*1/6 + 0*1/4 + 100600*1/4 + 196400*1/3 = 90616,

$\overline{R}_2$ = 49360*1/6+ 48300*1/4 +0*1/4 + 95800*1/3 = 52235,

$\overline{R}_3$ = 50440*1/6+98800*1/4+101000*1/4 + 0*1/3 = 65023.

Мінімальний середній ризик дорівнює $\overline{R}_2$ =52235 і відповідає стратегії компанії P_2.

Розглянута оптимізаційна задача є двокритеріальною, у ній кожне рішення має два критерії - середній прибуток і середній ризик.

Існує кілька способів постановки таких оптимізаційних задач. Розглянемо один з них у загальному вигляді.

Нехай Q - певна множина операцій. Кожна операція «q» має дві числові характеристики E(q) і R(q) (наприклад, ефективність і ризик) і різні операції обов'язково розрізняються хоча б однією характеристикою. При виборі найкращої операції бажано, щоб E було великим, а R малим.

Будемо говорити, що операція a домінує операцію b, і позначати $a>b$, якщо $E(a) > E(b)$ і $R(a) > R(b)$ і хоча б одна з цих нерівностей є строгою. При цьому операцію *a* називаюєть **домінуючою**, а операцію *b* - **домінуємою**. Ясно, що ні при якому розумному виборі найкращої операції домінуєма операція не може бути визнана такою. Отже, найкращу операцію треба шукати серед недомінуємих операцій. Множину цих операцій називають множиною Парето або множиною оптимальності за Парето.

На множині Парето кожна з характеристик E і R - однозначна функція іншої, тобто за характеристикою E можна визначити характеристику R і навпаки.

Щодо матричних ігор розподіл називають Парето-оптимальним, якщо становище жодного з гравців не можна поліпшити, не погіршуючи при цьому становища його партнера.

Продовжимо аналіз розглянутого прикладу. Кожне рішення ($\overline{R}_i$, $\overline{E}_i$) позначимо як точку на площині (рис. 11.1), одержимо три точки.

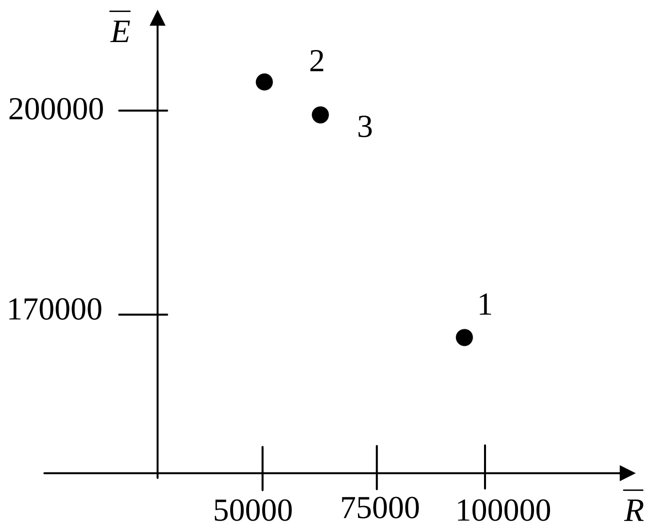


Рисунок 11.1– Множина операцій

Чим вище точка ($\overline{R}_i$, $\overline{E}_i$), тим більш дохідна операція, чим точка правіше, тим вона більш ризикова. Виходить, потрібно вибирати точку вище й лівіше. У розглянутому прикладі множина Парето складається тільки з одної другої операції.

Для знаходження кращої операції іноді застосовують зважуючу формулу, яка для операції E з характеристиками ($\overline{R}$, $\overline{E}$) дає одне число, на підставі якого й визначають кращу операцію. Наприклад, зважуюча формула має вигляд: $f(E) = 2\ \overline{E}- \ \overline{R}$. Тоді маємо:

$$f(E_1) = 2 *172550-906161 = 2544831,$$
$$f(E_2) = 2*210931-52235 = 369628,$$
$$f(E_3) = 2 *204810-65023 = 344596.$$

Звідси видно, що стратегія E_2 - краща.

Зважуюча формула виражає ставлення особи, що приймає рішення до доходу й ризику. Якщо ОПР застосовує розглянуту формулу, то вона згодна на збільшення ризику операції на дві одиниці, якщо дохід операції при цьому збільшується не менш чим на одну одиницю.

Контрольні запитання

1. У яких випадках для розв'язання ризикової ситуації використовують спеціальні критерії?
2. Поясніть, в чому полягає сутність критерію Вальда? Як він формулюється у разі пошуку мінімуму витрат?
3. Поясніть, в чому полягає сутність критеріїв оптимізму і песимізму? Як формулюються ці критерії у разі пошуку мінімуму витрат?

4. Поясніть, в чому полягає сутність критерію Севіджа?
5. Поясніть, в чому полягає сутність критерію Гурвіца?
6. Як на практиці відбирають розглянуті критерії?
7. У яких випадках використовують інтегральний критерій оптимальності? Поясніть його сутність.
8. Множину яких операцій називають множиною оптимальності за Парето?

ТЕМА 12
СИСТЕМА ПОКАЗНИКІВ КІЛЬКІСНОЇ ОЦІНКИ СТУПЕНЯ РИЗИКУ

Для ідентифікації ризику, як і інших елементів системи керування ним, велике значення має гарна інформаційна база, що складається із збору й обробки відповідної інформації. Відсутність відповідної інформації - важливий фактор будь-якого ризику.

Для оцінки ступеня ризику використовують якісний і кількісний аналіз. **Якісний аналіз** - це аналіз джерел і потенційних зон ризику, зумовлених його факторами. Тому якісний аналіз спирається на чітке виділення факторів, перелік яких є специфічним для кожного виду ризику. **Кількісний аналіз** ризику має на меті чисельно визначити, тобто формалізувати ступінь ризику. У кількісному аналізі можна виділити умовно кілька блоків: вибір критеріїв оцінки ступеня ризику; визначення припустимого рівня окремих видів ризику; визначення фактичного ступеня ризику на основі окремих методів; оцінка можливості збільшення або зниження ризику надалі.

Ризик розглядають як можливість втрат, що виникає внаслідок необхідності прийняття рішень в умовах невизначеності. Ступінь цієї можливості можна характеризувати різними показниками:

- імовірністю настання події;
- величиною відхилення від прогнозованого значення;
- дисперсією, математичним сподіванням, середнім квадратичним відхиленням, коефіцієнтом асиметрії, ексцесом, а також множиною інших математичних і статистичних показників.

Оскільки невизначеність може бути заданою різними її видами (імовірнісні розподіли, інтервальна невизначеність, суб'єктивні імовірності та ін.), а прояви ризику надзвичайно різноманітні, на практиці доводиться використовувати весь арсенал перелічених показників. У загальному випадку найчастіше використовують математичне сподівання M і середнє квадратичне відхилення σ які добре зарекомендували себе на практиці. Отже, рекомендується описувати ризик двома параметрами R (P; $П$).

12.1 Імовірнісна оцінка ризику

Ризик - категорія імовірнісна, тому в процесі оцінки невизначеності і кількісного визначення ризику використовують імовірнісні розрахунки. Кількісна оцінка імовірності настання окремих ризиків і того, у що вони можуть обійтися, дозволяє виділити найімовірніші за виникненням і вагомі за величиною втрат ризики, які будуть об'єктом подальшого аналізу для прийняття рішення про доцільність реалізації проекту. Оцінка імовірності також дозволяє усвідомити практичні можливості вибіркових досліджень і дати прогноз майбутніх дій.

Імовірнісні задачі характеризуються тим, що ефективність прийнятих рішень залежить не тільки від детермінованих факторів, але й від імовірностей випадкових факторів. Нехай відомий закон розподілу керованих факторів X у вигляді ряду розподілу випадкової величини X, що наведений у таблиці 12.1.

Таблиця 12.1 - Ряд розподілу випадкової величини X

x	x_1	x_2	…	x_n
P	P_1	P_2	…	P_n

де P_i - імовірність того, що керований фактор x_i, $i=\overline{1,n}$ прийняв те або інше значення.

Кожній парі (x_i, P_i) відповідає значення функції ефективності $E(x_i, P_i)$. Як показники ефективності зазвичай використовують математичне сподівання M, дисперсію D, середнє квадратичне відхилення σ, коефіцієнт варіації V та інші імовірнісні характеристики, значення яких у цьому випадку обчислюють за формулами

$$M=\sum x_i p_i\,,\quad D=\sigma^2=\sum(x_i-M)^2 P_i=\sum x_i^2 P_i - M^2\,,\quad V=\pm\frac{\sigma}{M}100\%. \quad (12.1)$$

Середнє M є узагальненою кількісною характеристикою фактору й не дозволяє прийняти рішення на користь будь-якого варіанта.

Середнє квадратичне відхилення σ є іменованою величиною й виміряється в тих самих одиницях, у яких виміряється ознака, що варіює. Дисперсія D і середнє квадратичне відхилення σ є мірами абсолютної мінливості.

Дисперсія не дає повної картини лінійних відхилень $\Delta X = X - M$, наочніших для оцінювання ризиків. Проте, значення дисперсії дозволяє встановити зв'язок між лінійним і квадратичним відхиленнями за допомогою відомої нерівності Чебишева.

Імовірність P того, що випадкова величина X відхиляється від свого математичного сподівання більш ніж на задане значення $\varepsilon > 0$, не перевершує її дисперсії, поділеної на ε^2, тобто

$$P\{|X-M|>\varepsilon\}\le\frac{D}{\varepsilon^2}. \quad (12.2)$$

Звідси видно, що незначному ризику за середнім квадратичним відхиленням відповідає малий ризик і за лінійними відхиленнями: точки X з великою імовірністю розташовуватимуться усередині ε - оточення середнього значення M.

При оцінці ризику операції значенням середнього квадратичного відхилення керованого фактору (наприклад, доходу) *X*, вводять позначення $r = \sigma$.

Якщо, наприклад, під *X* розуміти випадковий дохід *Q*, то M_Q є середнім доходом, або ефективністю, а середнє квадратичне відхилення σ_Q є оцінкою ризикованості, ризиком і позначається *r*.

Коефіцієнт варіації *V* є безрозмірною величиною. З його допомогою можна порівнювати мінливість ознак, виражених у різних одиницях виміру. Значення коефіцієнта варіації лежать у межах від 0 до 100%. Чим більше коефіцієнт *V*, тим сильніше мінливість. На практиці використовують якісну оцінку різних значень коефіцієнта варіації: до 10% - слабка мінливість; 10-25% - помірна мінливість, понад 25% - висока мінливість.

За допомогою оцінки ризику на основі розрахунку дисперсії, стандартного відхилення й коефіцієнта варіації можна оцінити ризик конкретної операції або підприємницької фірми в цілому за певний проміжок часу.

Перевагою даного методу оцінки підприємницького ризику є простота математичних розрахунків, а недоліком - необхідність великої кількості вихідних даних.

Розглянемо приклад. Порівняємо за ризиком вкладення капіталу в акції трьох типів *A*, *B*, *C*, якщо для кожного з них відомі імовірності значень прибутковості, наведені в таблиці 12.2.

Таблиця 12.2 - Імовірності значень прибутковості акцій

Тип акцій	**Ситуація 1**		**Ситуація 2**	
	імовірність	**прибутковість**	**імовірність**	**прибутковість**
A	0,5	20%	0,5	10%
B	0,99	15,1%	0,01	5,1%
C	0,7	13%	0,3	7%

Визначимо для акції *A* середню прибутковість:

$$M_A = 20\% \cdot 0{,}5 + 10\% \cdot 0{,}5 = 15\%;$$

дисперсію

$$D_A = (20\% - 15\%)^2 \cdot 0{,}5 + (10\% - 15\%)^2 \cdot 0{,}5 = 25;$$

середнє квадратичне відхилення

$$\sigma_A = \sqrt{D_A} = 5\%;$$

коефіцієнт варіації

$$V_A = \frac{\sigma_A}{E_A}100\% = \frac{5}{15}100\% = 33{,}3\% .$$

Для акції B аналогічно одержимо:

$$M_B = 15{,}1\% \cdot 0{,}99 + 5{,}1\% \cdot 0{,}01 = 15\%;$$

$$D_B = (15{,}1\% - 15\%)^2 \cdot 0{,}99 + (5{,}1\% - 15\%)^2 \cdot 0{,}01 = 0{,}99;$$

$$\sigma_B = 0{,}995\%, \; V_B = \frac{0{,}995}{15}100\% = 6{,}63\% .$$

Для акції C:

$$M_C = 13\% \cdot 0{,}7 + 7\% \cdot 0{,}3 = 11{,}2\%;$$

$$D_C = (13\% - 11{,}2\%)^2 \cdot 0{,}7 + (7\% - 11{,}2\%)^2 \cdot 0{,}3 = 7{,}56;$$

$$\sigma_3 = 2{,}75\%,\ V_C = \frac{2{,}75}{11{,}2}100\% = 24{,}6\%.$$

Оскільки найменше значення коефіцієнта варіації в акцій B, то і вкладення в цю акцію найкращі, причому й $\sigma_B = r = 0{,}995\%$ найменші.

Особливий варіант ризику пов'язаний з розоренням. Так називають імовірність настільки великих втрат ($x < M$), які неможливо компенсувати.

Наприклад, випадковий дохід операції Q має наступний ряд розподілу:

Таблиця 12.3 - Ряд розподілу операції Q

Q_i	-60	-40	-30	80
P_i	0,1	0,2	0,5	0,2

Втрати, що дорівнюють або перевищуючі 30 ведуть до розорення. Тоді імовірність виникнення ризику розорення в результаті цієї операції

$$0{,}1 + 0{,}2 + 0{,}5 = 0{,}8.$$

Якщо імовірність дуже мала, то нею зневажують (імовірність розорення відмінна від нуля майже в будь-якій операції).

Визначимо імовірнісну міру ризику розорення.

Розглянемо приклад. Нехай на ринку можуть виникнути тільки два результати й на кожний з них акції A і B відгукуються невипадковим способом. Імовірності цих наслідків і відповідних ним значень прибутковості приведені в таблиці 12.4.

Таблиця 12.4 - Імовірності наслідків і значення прибутковості акцій

Тип акцій	**Результат 1**		**Результат 2**	
	імовірність	**прибутковість**	**імовірність**	**прибутковість**
A	0,3	6%	0,7	2%
B	0,2	-1%	0,8	4,25%

Середня прибутковість акцій A і B:

$M_A = 6\% \cdot 0{,}3 + 2\% \cdot 0{,}7 = 3{,}2\%$, $M_B = -1\% \cdot 0{,}2 + 4{,}25\% \cdot 0{,}8 = 3{,}2\%$,

дисперсії відповідно дорівнюють:

$D_A = (6\%- 3{,}2\%)^2 \cdot 0{,}3 + (2\% - 3{,}2\%)^2 \cdot 0{,}7 = 3{,}35$; $\sigma_A = r_A = 1{,}83$;

$D_B = (-1\% - 3{,}2\%)^2 \cdot 0{,}2 + (4{,}25\% - 3{,}2\%)^2 \cdot 0{,}8 = 3{,}41$; $\sigma_B = r_B = 1{,}85$.

Припустимо тепер, що інвестор взяв гроші в борг під 2,5%. Ставка відсотка за кредитом нижче за очікувану прибутковість по акціях, які будуть придбані на позичені гроші, тому дії інвестора цілком розумні.

Проте, якщо інвестор вклав гроші в акції A, то при результаті 1 він виграє (6% - 2,5%) = 3,5%, а при результаті 2 програє (2% - 2,5%) = -0,5%, причому з імовірністю $P_2 = 0{,}7$. Навпроти, якщо він вкладе гроші в актив B, то програш йому загрожує з імовірністю $P_1 = 0{,}2$ у першій ситуації (результат 1), коли він втратить (-1%-2,5%) = -3,5%.

Підрахуємо очікувані втрати (*П*) при покупці акцій *А* і *В* відповідно: $ПА$ = 0,5% • 0,7 = 0,35%; $П_В$ = 3,5% • 0,2 = 0,7%. Як бачимо, у першому випадку вони менші, але ризик втрат при придбанні акцій *А* більший (0,7 > 0,2). Це повинне схилити обережного вкладника на користь акцій *В*. У свою чергу, очікуваний ризик $П_А < П_В$ схиляє його до вибору на користь акцій *А*. Як діяти в подібній ситуації? Це залежить від індивідуальних переваг, що виражаються в тому числі функцією корисності інвестора.

У розглянутих імовірнісних методах використовують поняття: ризик (функція ризику), втрати (функція втрат), рішення (функція рішення), функції розподілу за певних умов. Випадок прийняття рішення в умовах ризику перебуває між визначеними (детермінованими) умовами і невизначеністю. При цьому можна оцінити імовірність виникнення кожної випадкової події. Широко використовують при таких обставинах критерій передбачуваного виграшу.

Передбачуваний виграш розраховують для кожної альтернативи, після чого відбирають альтернативу з найвищим показником. Передбачуваний виграш визначають як суму значень виграшу кожної альтернативи, у якій кожний доданок зважується з погляду імовірності відповідної альтернативи.

У випадку стохастичної невизначеності, коли некерованим факторам (станам природи) поставлені у відповідність імовірності, задані експертно або обчислені, рішення зазвичай приймають на основі критерію максимуму очікуваного середнього виграшу або мінімуму середнього ризику.

12.2 Статистичні оцінки показників ризику

Будь-які економічні дані є кількісними характеристиками будь-яких економічних об'єктів. Вони формуються під дією множини факторів, не всі з яких доступні зовнішньому контролю. Стохастична природа економічних даних зумовлює необхідність застосування спеціальних адекватних їм статистичних методів для обробки та аналізу.

Стосовно економічних задач методи математичної статистики збігаються до систематизації, обробці та використання статистичних даних для наукових і практичних висновків. Метод дослідження, що спирається на розгляд статистичних даних про ті або інші сукупності об'єктів, називають статистичним. Основним елементом економічного дослідження є аналіз і побудова взаємозв'язків економічних змінних. Вивчення таких взаємозв'язків ускладнене тим, що вони не є строгими функціональними залежностями. Буває досить важко виявити всі основні фактори, що впливають на досліджувану змінну (наприклад, прибуток, ризик), багато таких взаємодій є випадковими, носять невизначений характер, і число статистичних спостережень є обмеженим. У цих умовах математична статистика дозволяє будувати економічні моделі й оцінювати їх параметри, перевіряти гіпотези про властивості економічних показників і форми їх зв'язку, що є основою для прийняття обґрунтованих економічних рішень. Теорія імовірностей відіграє важливу роль при статистичних дослідженнях випадкових явищ. Тут повною мірою знаходять застосування такі засновані на теорії імовіностей розділи математичної статистики, як статистична перевірка гіпотез, статистичне оцінювання розподілів імовірностей та ін.

Для оцінки величини ризику використовують статистичну вибірку, представлену варіаційним рядом, що наведений у таблиці 12.5.

Таблиця 12.5 - Варіаційний ряд

x_i	x_1	x_2	…	x_n
n_i	n_1	n_2	…	n_n

Для оцінки середнього значення та мінливості досліджуваної результативної ознаки використовують формули:

$$\bar{x} = \frac{\sum x_i n_i}{\sum n_i}; \quad \sigma^2 = \frac{\sum \left(x_i - \bar{x}\right)^2 n_i}{\sum n_i}; \quad V = \pm \frac{\sigma}{\bar{x}} 100\%, \tag{12.3}$$

де x_i - середнє значення фактору для кожного спостереження (і = 1,2,...);

n_i - число спостережень (частота) значення x_i;

$\bar{x}$ - середнє варіаційного ряду;

σ^2 — дисперсія;

V - коефіцієнт варіації.

Отже, мінливість результативної ознаки виражається статистичними оцінками дисперсії σ^2, середнього квадратичного відхилення σ і коефіцієнта варіації V.

Розглянуті вище показники ризику передбачається застосовувати до нормального розподілу імовірностей. Цей розподіл широко використовують при аналізі ризиків господарських операцій, тому що його найважливіші властивості (симетричність розподілу щодо середнього, незначна імовірність великих відхилень випадкової величини від центра її розподілу, правило трьох сигм) дозволяє істотно спростити аналіз. Однак не всі господарські операції припускають нормальний розподіл доходів. Наприклад, розподіли імовірностей одержання доходів від операцій з похідними фінансовими інструментами (опціонами й ф'ючерсами) часто характеризується асиметрією (скосом) щодо математичного сподівання випадкової величини (рис. 12.1).

Так, наприклад, опціон на покупку цінного папера дозволяє його власникові дістати прибуток у випадку позитивної прибутковості й одночасно уникнути збитків у випадку негативної, тобто по суті, опціон відсікає розподіл прибутковості в точці, де починаються втрати.

У подібних випадках використання для аналізу тільки двох параметрів (середньої й стандартного відхилення) може приводити до невірних висновків. Стандартне відхилення неадекватно характеризує ризик при зміщених розподілах, оскільки ігнорується те, що більша частина мінливості припадає на «гарну» (праву) або «погану» (ліву) сторону очікуваної прибутковості. Тому при аналізі асиметричних розподілів використовують додатковий параметр – коефіцієнт асиметрії (скосу). Він є нормованою величиною третього центрального моменту μ_3 і визначається за формулою:

$$Sk = \frac{\mu_3}{\sigma^3} = \frac{\sum (x_i - \bar{x})^3}{n\sigma^3}. \tag{12.4}$$

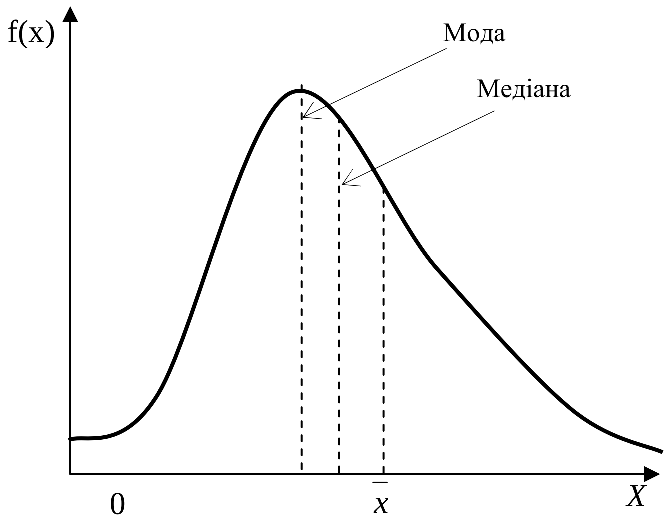


Рисунок 12.1 – Асиметричне розподілення імовірностей

Економічний зміст коефіцієнта асиметрії полягає в наступному. Якщо коефіцієнт має позитивне значення (позитивний скіс), то найвищі доходи (правий «хвіст») вважають імовірнішими, ніж низькі і навпаки.

Коефіцієнт асиметрії може також використовуватися для наближеної перевірки гіпотези про нормальний розподіл випадкової величини. Його значення в цьому випадку повинне дорівнювати 0.

У ряді випадків зміщений вправо розподіл можна звести до нормального додатком одиниці до очікуваної величини прибутковості й наступним обчисленням натурального логарифму отриманого значення. Такий розподіл називають логнормальним. Його використовують у фінансовому аналізі поряд з нормальним.

Певні симетричні розподіли можуть характеризуватися четвертим нормованим центральним моментом - ексцесом ε:

$$\varepsilon_x = \frac{\mu_4}{\sigma^4} - 3 = \frac{\sum (x_i - \bar{x})^4}{n\sigma^4} - 3. \qquad (12.5)$$

Якщо значення ексцесу перевищує 0, крива розподілу гостріша, ніж нормальна крива й навпаки.

Економічний зміст ексцесу полягає в наступному. Якщо дві операції мають симетричні розподіли доходів і однакові середні, менш ризикованою вважають операцію з більшим ексцесом. Для нормального розподілу ексцес дорівнює 0.

12.3 Аналіз ризиків господарських операцій на підставі нормального розподілу

Один з імовірнісних підходів до дослідження ризиків заснований на припущенні, що більшість показників господарської діяльності (прибуток, дохід та ін.) як випадкові величини підпорядковуються нормальному закону розподілу. Цей закон має місце, коли досліджувана випадкова величина є результатом спільного впливу великої кількості незалежних факторів, жоден з яких не робить на неї переважного впливу.

Нормальний розподіл є основним елементом більшості систем управління ризиком і дає більше важливої інформації, чим просто оцінки параметрів вибірки. На практиці для перевірки припущення про нормальний розподіл досліджуваної сукупності випадкових факторів застосовують критерії згоди, що встановлюють відповідність між емпіричним і теоретичним розподілами.

Щільність імовірності нормального розподілу має вигляд:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{(x-\bar{x})^2}{2\sigma^2}} . \qquad (12.6)$$

З курсу теорії імовірностей відомо, що влучення випадкової величини X у заданий інтервал (α, β) визначають за формулою

$$P\{\alpha < X < \beta\} = \int_{\alpha}^{\beta} f(t)dt = \Phi\left(\frac{\beta - \bar{x}}{\sigma}\right) - \Phi\left(\frac{\alpha - \bar{x}}{\sigma}\right), \qquad (12.7)$$

де $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_0^x e^{\frac{(x-\bar{x})^2}{2\sigma^2}} dx$ - інтеграл імовірностей або функція Лапласа, її значення залежно від параметра x приводяться в спеціальних таблицях. Інтеграл імовірностей - функція непарна, вона змінюється від 0 до 0,5.

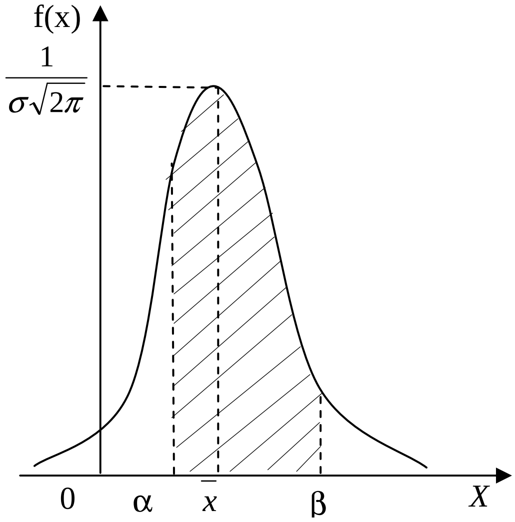


Рисунок 12.2 – Нормальна крива

Якщо припустити, що очікуване значення результативної ознаки X (прибуток, втрати та ін.) належить інтервалу (α, β), то імовірність цієї події визначається за формулою (12.7). Нехай вона дорівнює P_1. На графіку щільності імовірностей (рис. 12.2) заштрихована площа чисельно дорівнює P_1. Тоді імовірність влучення досліджуваної результативної ознаки за межі інтервалу (α, β) дорівнює $P_2 = 1 - P_1$.

Середнє значення $\bar{x}$ визначає центр розподілу. Середнє квадратичне відхилення σ - розкид X від центра розподілу. Середнє значення $\bar{x}$ впливає на положення графіка на осі X, а значення σ змінює розмах кривої.

На практиці часто як характеристику розсіювання використовують не середнє квадратичне відхилення σ, а величину, називану імовірним відхиленням (інакше - «серединним відхиленням» або «серединною помилкою»).

Серединним відхиленням називають половина довжини ділянки, симетричної щодо центра розсіювання, імовірність влучення в яку дорівнює 0,5. Геометрично серединне відхилення E є половиною довжини ділянки осі абсцис, симетричної щодо центра розсіювання, на яку спирається половина площі кривої розподілу (рис. 12.3).

Пояснимо зміст терміна «серединне відхилення». Імовірність того, що величина X відхилиться від центра розсіювання $\bar{x}$ менше чим на E, за визначенням дорівнює

$$P\{|X - \bar{x}| < E\} = \frac{1}{2}. \qquad (12.8)$$

Імовірність того, що це відхилення перевищуватиме E, також дорівнює 1/2

$$P\{|X - \bar{x}| > E\} = \frac{1}{2}.$$

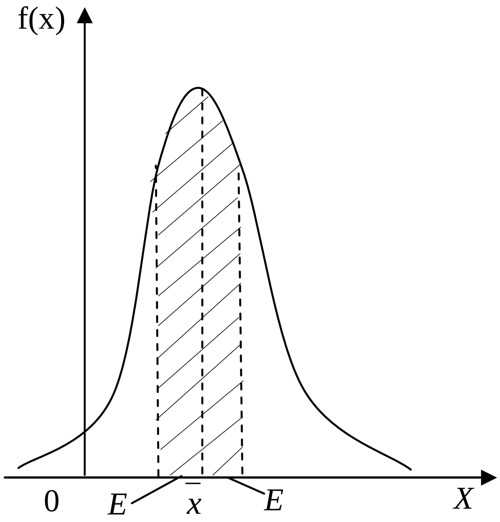


Рисунок 12.3 – Серединне відхилення

Отже, при великій кількості дослідів у середньому половина значень випадкової величини X відхилятиметься від $\bar{x}$ більше чим на E, а половина - менше, звідси й термін «серединне відхилення».

З курсу теорії імовірностей відомо, що імовірність того, що відхилення випадкової величини X від середнього значення $\bar{x}$ за абсолютною величиною не перевищить числа $\varepsilon=\sigma t$, визначається співвідношенням

$$P\left\{\left|X-\bar{x}\right|<\varepsilon\right\}=2\Phi\left(\frac{\sigma t}{\sigma}\right)=2\Phi(t). \quad (12.9)$$

Очевидно, серединне відхилення як характеристика розсіювання повинне перебувати в прямій залежності від середнього квадратичного відхилення σ. Встановимо цю залежність, для чого обчислимо імовірність події $|X-\bar{x}|<E$ у рівнянні (12.8) за формулою (12.9)

$$P\left\{\left|X-\bar{x}\right|<E\right\}=2\Phi\left(\frac{E}{\sigma}\right)=\frac{1}{2}, \quad (12.10)$$

звідки одержимо $\frac{E}{\sigma}=0{,}0987; \quad E\approx 0{,}1\sigma$.

Формули (12.7), (12.9) і (12.10) застосовують на практиці для обчислення імовірності влучення X у заданий інтервал.

За визначенням серединного відхилення імовірність влучення на ділянку довжини E, що примикає до центра розсіювання, дорівнює 0,25. Отже, можна вважати практично достовірним те, що випадкова величина, підлегла нормальному закону, відхиляється від центра розсіювання не більше чим на чотири серединних відхилення.

12.4 Крива ризиків

У процесі оцінки ризику прийнято розрізняти зони ризику залежно від рівня можливих втрат. У таблиці 12.6 приведено емпіричну шкалу ризику.

Таблиця 12.6 - Емпірична шкала припустимого рівня ризику

№ з/п	**Імовірність небажаного результату (величина ризику)**	**Найменування градацій ризику**
1	0,0-0,1	мінімальний
2	0,1-0,3	малий
3	0,3-0,4	середній
4	0,4-0,6	високий
5	0,6-0,8	максимальний
6	0,8-1,0	критичний

Використовуючи співвідношення (12.9) і вибираючи імовірності з таблиці 12.3, за таблицями функції Лапласа $\Phi(t)$ знаходимо відповідні значення параметра t (табл. 12.7).

Таблиця 12.7 - Значення імовірностей і параметра t

P	0	0,1	0,3	0,4	0,5	0,6	0,6826	0,8	0,9544	0,9973
t	0	0,126	0,386	0,524	0,674	0,842	1	1,281	2	3

Нанесемо значення $\varepsilon=\sigma t$ на графік нормальної кривої вправо від $\bar{x}$ і одержимо зони ризику (рис. 12.4). Перша точка кривої визначає імовірність нульових втрат. Друга точка відповідає «нормальному», «розумному» ризику, при якому рекомендується приймати звичайні підприємницькі рішення. Зона прийнятного ризику характеризується рівнем втрат, що не перевищує розміру чистого прибутку. Третя точка характеризується величиною можливих втрат, що дорівнюють очікуваному прибутку, тобто повної втрати прибутку. Четверта точка відповідає величині втрат, що дорівнюють розрахунковій виручці. П'ята точка характеризується втратами, що дорівнюють майновому стану підприємця.

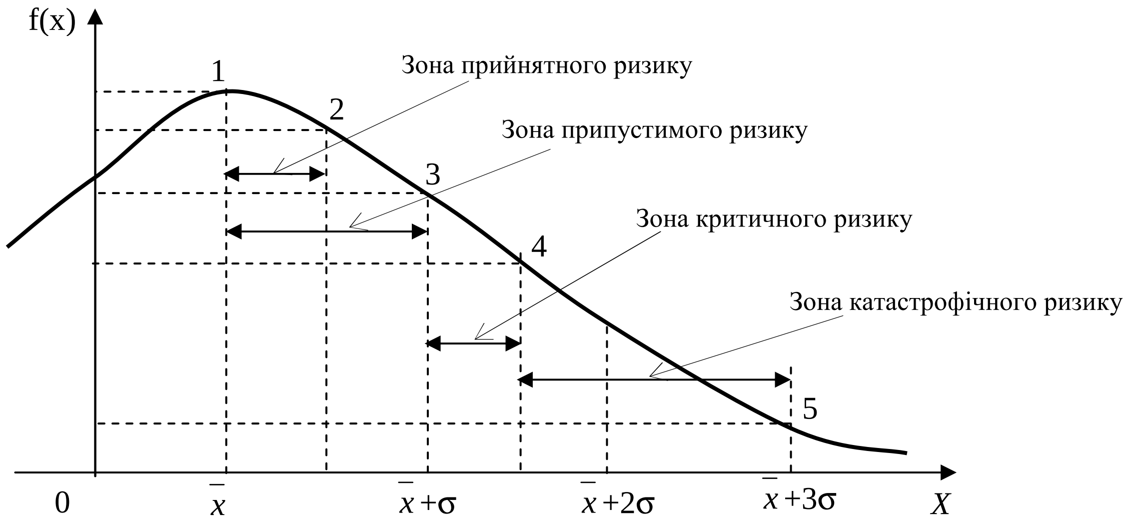


Рисунок 12.4 – Крива ризиків

Розглянутим точкам ризику відповідають наступні значення імовірностей:

$$P_1 \le 0{,}1;\ P_2 = 0{,}25;\ P_3 = 0{,}4;\ P_4 = 0{,}75;\ P_5 > 0{,}75.$$

Знання граничних значень імовірностей виникнення припустимого P_D, критичного $P_{кр}$ і катастрофічного $P_{кат}$ ризиків дозволяє сформулювати самі загальні умови прийнятності аналізованого виду підприємництва:

- показник припустимого ризику не повинен перевищувати граничного значення, тобто $P_3 < P_D$;

- показник критичного ризику повинен бути меншим за граничну величину, тобто $P_4 < P_{кр}$;

- показник катастрофічного ризику не повинен перевищувати граничного рівня, тобто $P_5 < P_{кр}$.

12.5 Оцінка ризиків за допомогою довірчих інтервалів

Якщо результати економічної діяльності підлягають нормальному закону розподілу імовірностей, то в цьому випадку має місце, так зване, правило трьох сигм, що у більш широкій постановці дозволяє встановити область можливих значень випадкової величини X як

$$M-t\sigma<X<M+t\sigma, \qquad (12.11)$$

де величина t характеризує довірчий інтервал випадкової величини X. При $t = 1$ з імовірністю 0,6826 (або в 68% випадків) можна стверджувати, що значення випадкової величини лежить у межах $M \pm \sigma$; при $t =2$ з імовірністю 0,9544 можна стверджувати, що $X \in (M-2\sigma, M+ 2\sigma)$; і при $t = 3$ імовірність того, що значення випадкової величини $X \in (M-3\sigma, M+3\sigma)$ становить 0,9973.

Імовірнісна постановка задачі вибору оптимальних рішень в економіці адекватніше відображає реальні ситуації. Тому застосування імовірнісних моделей у багатьох випадках дозволяє зменшити ризик при виборі найефективніших рішень. Однак застосування зазначених моделей пов'язане з необхідністю визначення імовірнісних характеристик аналізованих процесів, що істотно ускладнює рішення розглянутих задач. У багатьох випадках імовірнісний розподіл економічних показників буває невідомим. Тому виникає необхідність визначення кращих альтернатив за умови, що імовірнісні характеристики економічних показників є невідомими.

В умовах повної невизначеності, коли імовірності розглянутих ситуацій невідомі, можна користуватися правилом Лапласа, яке полягає в тому, що всі невідомі імовірності P_i вважають рівними. Після цього вибір ефективного рішення можна приймати або за правилом максимізації середнього очікуваного виграшу або за правилом мінімізації середнього ризику. Подібний критерій прийняття рішення можна назвати принципом недостатнього обґрунтування Лапласа.

12.6 Вибір розподілу випадкової величини

Нормальний розподіл використовують, коли неможливо точно визначити імовірність того, що безперервна випадкова величина приймає якесь конкретне значення. Нормальний розподіл припускає, що варіанти прогнозованого параметра тяжіють до середнього значення. Значення параметра, які істотно відрізняються від середнього, тобто перебувають в «хвостах» розподілу, мають малу імовірність. Така природа нормального розподілу.

Трикутний розподіл є сурогатом нормального й припускає лінійно наростаючий в міру наближення до моди розподіл.

Трапецієподібний розподіл припускає наявність інтервалу значень із найбільшою імовірністю реалізації.

Рівномірний розподіл вибирають, коли припускають, що всі варіанти прогнозованого показника мають однакову імовірність реалізації.

Якщо випадкова величина дискретна, а не безперервна, застосовують біноміальний розподіл і розподіл Пуассона. Розподіл Пуассона застосовують при виконанні наступних умов:

- кожний малий інтервал часу може розглядатися як дослід, результатом якого є одне з двох: або «успіх», або його відсутність – «промах». Інтервали настільки малі, що може бути тільки один «успіх» в одному інтервалі, імовірність якого мала й незмінна.

- число «успіхів» в одному великому інтервалі не залежить від їх числа в іншому, тобто «успіхи» безладно розкидані по часових проміжках.

- середнє число «успіхів» постійне протягом усього розглянутого інтервалу часу.

За певних умов розподіл Пуассона можна використовувати як апроксимацію біноміального розподілу, що особливо зручно коли застосування біноміального розподілу вимагає складних розрахунків. Апроксимація гарантує прийнятні результати при виконанні наступних умов:

- кількість дослідів велика, понад тридцять ($n \geq 30$);

- імовірність «успіху» у кожному досліді мала, менша за 0,1 ($p \leq 0{,}1$). Якщо імовірність «успіху» велика, то для заміни можна використати нормальний розподіл.

- передбачуване середнє число «успіхів» менше за 5 ($np \leq 5$).

Біноміальний розподіл також можна апроксимувати нормальним розподілом з «виправленням на безперервність», тобто роблячи допущення, що, наприклад, значення дискретної випадкової величини 2 є значенням безперервної випадкової величини на проміжку від 1,5 до 2,5. Оптимальна апроксимація досягається при виконанні наступних умов: $n \geq 30$; $np \leq 5$, а імовірність «успіху» $p \leq 0{,}1$ (оптимальне значення $p = 0{,}5$).

Слід зазначити, що на практиці крім статистичних критеріїв використовують й інші показники виміру ризику: величину упущеної вигоди, недоотриманий дохід та інші, що розраховують, як правило, у грошових одиницях. Безумовно, такі показники мають право на існування, більше того, вони найчастіше є простішими й зрозумілішими ніж статистичні критерії, однак для адекватного опису ризику вони повинні враховувати і його імовірнісну оцінку. Зокрема, дисперсію, средньоквадратичне відхилення і коефіцієнт варіації.

12.7 Імітаційне моделювання

Методи імітаційного моделювання дозволяють зібрати необхідну інформацію про поведінку системи шляхом створення її комп'ютеризованої моделі. Імітаційне моделювання не вирішує оптимізаційних задач, а є технікою оцінки значень функціональних характеристик системи, що моделюють. Методи імітаційного моделювання знаходять широке застосування в економічних і комерційних задачах, включаючи оцінку поведінки споживача, визначення цін, економічне прогнозування діяльності фірм, у соціальних задачах та ін.

Попередником сучасного імітаційного моделювання вважають метод Монте-Карло, основна ідея якого перебуває у використанні вибірки випадкових чисел для одержання імовірнісних або детермінованих оцінок будь-яких величин. Імітація є випадковим експериментом, тому будь-який результат імітаційного моделювання піддається експериментальним помилкам і підлягає статистичній перевірці. Для будь-якого експерименту також важливим є питання, яким повинен бути обсяг вибірки n і число реалізацій досліджуваної випадкової величини N.

Відмінність сучасних імітаційних моделей від методу Монте-Карло полягає в тому, що імітаційна модель зазвичай пов'язана з вивченням реально існуючої системи, поведінка якої є функцією часу. Існує два типи імітаційних моделей. **Безперервні моделі** використовують для систем, поведінка яких змінюється в часі безупинно. Безперервні імітаційні моделі зазвичай представляються у вигляді різницево-диференційних рівнянь, які описують взаємодію між різними елементами системи. **Дискретні моделі** описують системи, поведінка яких змінюється тільки в певні моменти часу. Типовим прикладом такої моделі є черга, що є системою, зміни в якій відбуваються лише тоді, коли клієнт надходить у чергу або залишає систему після обслуговування. Це означає, що в будь-якій дискретній імітаційній моделі є дві головних події, за якими необхідно досліджувати систему. В імітаційній моделі події, пов'язані із прибуттям, визначаються часом між надходженнями клієнтів, а події, пов'язані з їх доглядом, - часом обслуговування.

Випадковість в імітаційних моделях виникає тоді, коли інтервал часу t між однорідними подіями є випадковим. Відомий ряд методів одержання послідовних випадкових значень $t=t_1, t_2, \ldots$, що мають заданий розподіл імовірностей $f(x)$. Розглянемо два з них: метод зворотних функцій і метод згорток.

Обидва методи засновані на використанні незалежних однаково розподілених випадкових чисел, що мають рівномірний розподіл на інтервалі [0,1]. Метод зворотних функцій використовують для безперервних розподілів, наприклад для експонентного або рівномірного. Метод згорток використовують в складніших ситуаціях, наприклад, при генеруванні випадкових чисел, що мають нормальний розподіл або розподіл Пуассона.

Метод зворотних функцій вимагає виконання наступних дій. Спочатку генерується випадкове число R з інтервалу [0,1], потім обчислюється шукане випадкове число $x=F^{-1}(R)$, де F^{-1}– функція, зворотна до функції розподілу $F(x)=P\{X<x\}$ (рис. 12.5). Метод заснований на тому, що якщо функцію розподілу $F(x)$ розглядати як випадкову величину, то вона розподілена рівномірно на інтервалі [0,1].

Розглянемо приклад. Нехай час появи клієнтів розподілено за експонентним законом з параметром $\lambda=4$. Функція розподілу має вигляд $F(t)=1-e^{-\lambda t}=1-e^{-4t}$. Врахуємо, що $F(t)=R$, одержимо $t=\frac{1}{\lambda}\ln(1-R)$. Оскільки R - випадкове число з інтервалу [0,1] і $(1-R)$ також випадкове число з того самого інтервалу, можна замінити $(1-R)$ на R. Нехай $R=0,9$, тоді одержимо одне конкретне значення інтервалу часу між клієнтами $t=\frac{1}{4}\ln(1-0,9)=0,577$.

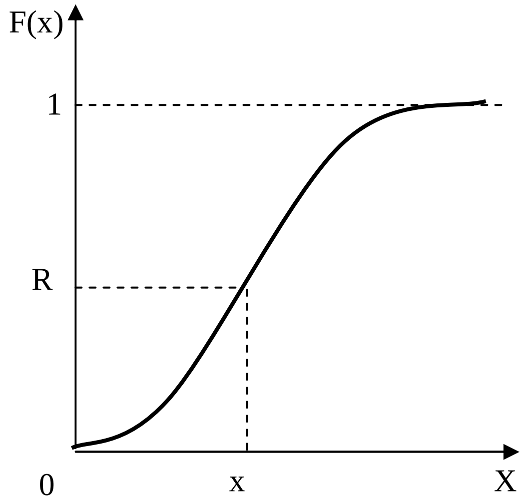


Рисунок 12.5 – Метод зворотних функцій

Значення R – мають обиратися випадково з інтервалу [0,1] і підлягати рівномірному розподілу.

Основна ідея методу згорток полягає в тому, щоб виразити шукану випадкову величину у вигляді суми інших випадкових величин, для яких легко отримати реалізації випадкових значень. Для одержання значень, що відповідають нормальному розподілу з математичним сподіванням M і стандартним відхиленням σ використовують центральну граничну теорему. Нагадаємо, що суть її зводиться до того, що сума n однаково розподілених випадкових величин прагне до нормального розподілу при нескінченому збільшенні n. Нехай $x=R_1+R_2+\ldots+R_n$, де R_1, R_2, …, R_n – випадкові числа, рівномірно розподілені в інтервалі [0,1]. Відповідно до центральної граничної теореми випадкова величина x є асимптотично нормальною величиною із середнім $n/2$ і дисперсією $n/12$. Тоді випадкова величина X з математичним сподіванням M і стандартним відхиленням σ визначається за формулою

$$X = M + \sigma\left(\frac{x - \frac{n}{2}}{\sqrt{\frac{n}{12}}}\right).$$

У практичних розрахунках для зручності зазвичай приймають n=12, тоді $X = M + \sigma(x - 6)$.

Імітаційне моделювання є статистичним експериментом. Його результати повинні ґрунтуватися на відповідних статистичних перевірках з використанням, наприклад, довірчих інтервалів і методів перевірки гіпотез. Для цього спостереження повинні задовольняти наступним вимогам: мати стаціонарні розподіли, тобто такі, що не змінюються під час проведення експерименту, підлягати нормальному розподілу й бути незалежними.

Характер імітаційних обчислень стимулює створення спеціалізованих мов програмування. У цей час на ринку програмних продуктів для моделювання домінують комерційні пакети Arena, AweSim і GPSS/H, які мають розвинений інтерфейс, що спрощує процес створення імітаційних моделей.

Контрольні запитання

1. В чому полягає якісний аналіз ризику і в чому кількісний? Охарактеризуйте різницю.
2. Чим відрізняються методи кількісної оцінки ризику на підставі теорії імовірностей і на підставі математичної статистики?
3. Поясніть, як середнє квадратичне відхилення σ використовують для оцінки ризику?

4. Поясніть, як використовують для оцінки ризику коефіцієнт асиметрії і ексцес?
5. Які переваги дає припущення про нормальний розподіл варіативної ознаки?
6. Поясніть, для чого використовують серединне відхилення і як воно пов’язане з середнім квадратичним відхиленням?
7. Які зони ризику розрізняють на кривій ризику? До якого закону розподілу належать ці зони?
8. В чому полягає принцип недостатнього обґрунтування Лапласа?
9. Для чого використовують метод імітаційного моделювання? У чому відмінність безперервних і дискретних імітаційних моделей?
10. Для чого призначений метод зворотних функцій? Поясніть суть методу.
11. Для чого призначений метод згорток? Поясніть суть методу.
12. Яка відмінність між прийняттям рішень в умовах визначеності, в умовах невизначеності й в умовах ризику?

ЗМ 3 Економетричні моделі

ТЕМА 13
ПРИНЦИПИ ПОБУДОВИ ЕКОНОМЕТРИЧНИХ МОДЕЛЕЙ

13.1 Роль економетричних досліджень в економіці

Економетричне моделювання реальних соціально-економічних процесів і систем зазвичай переслідує одну з двох кінцевих прикладних цілей: прогноз економічних і соціально-економічних показників, що характеризують стан і розвиток аналізованої системи, імітацію різних можливих сценаріїв соціально-економічного розвитку аналізованої системи (різноманітні сценарні розрахунки, ситуаційне моделювання).

При постановці задач економетричного моделювання треба визначити їх ієрархічний рівень і профіль. Аналізовані задачі можуть належати до макро- (країна, міждержавний аналіз), мезо- (регіони усередині країни) і мікро- (підприємства, фірми, родини) рівнів і бути спрямованими на вирішення питань різного профілю інвестиційної, фінансової або соціальної політики, ціноутворення, розподільних стосунків та ін.

Економетрія поєднує сукупність методів і моделей, що дозволяють на базі економічної теорії, економічної статистики і математико-статистичного апарата надавати **кількісні** вираження **якісним** залежностям. Основні результати економічної теорії мають якісний характер, а економетрія вносить до них емпіричний зміст. Математична економіка виражає економічні закони у вигляді математичних співвідношень, а економетрія здійснює дослідну перевірку цих законів. Економічна статистика дає інформаційне забезпечення досліджуваного процесу у вигляді вихідних (оброблених) статистичних даних і економічних показників, а економетрія, використовуючи традиційні математико-статистичні і спеціально розроблені методи, аналізує кількісні взаємозв'язки між цими показниками.

Багато базових понять економетрії мають два визначення - економічне і математичне. Подібна подвійність має місце й у формулюванні результатів. Економічна складова економетрії, безумовно, є первинною. Саме економіка визначає постановку задачі і вихідні передумови, а результат, формований математичною мовою, складає інтерес в тому випадку, якщо вдається його економічна інтерпретація. У той самий час велика кількість економетричних результатів має характер математичних стверджень (теорем).

Загальним моментом для будь-якої економетричної моделі є розбивка залежної змінної на дві частини - **пояснену** і **випадкову**. Сформулюємо задачу моделювання в самому загальному вигляді: на підставі експериментальних даних визначити пояснену частину і, розглядаючи випадкову складову як випадкову величину, отримати (можливо, після певних припущень) оцінки параметрів її розподілу. Отже, економетрична модель має вигляд, що наведений на рисунку 13.1.

Спостережене значення залежної змінної	=	Пояснена частина, залежна від значень змінних, що пояснюють	+	Випадкова складова

Рисунок 13.1 – Структура економетричної моделі

Нехай отримано наступний вираз для поясненої частини змінної Y:

$$y = b + b_1x_1 + b_2x_2.$$

Очевидно, що він дає уявлення про те, як саме формується розглянута економічна змінна y і можливість виявити вплив на неї кожної з пояснюючих змінних x_1 і x_2. В цьому випадку пояснювальна змінна y дорівнює b при x_1=0 і x_2= 0. За рахунок збільшення x_1 на одну одиницю вона збільшується на b_1, а за рахунок збільшення x_2 на одну одиницю зростає на b_2. Найбільш важливим є те, що отриманий вираз дозволяє прогнозувати значення поясненої частини змінної Y, якщо відомі його параметри b, b_1 і b_2.

13.2 Етапи економетричного моделювання

Прийнято виділяти шість основних етапів економетричного моделювання: постановочний, апріорний, етап параметризації, інформаційний, етапи ідентифікації і верифікації моделі.

Зупинимося докладніше на кожному з цих етапів і розглянемо проблеми, пов'язані з їх реалізацією.

1-й етап (постановочний). На даному етапі формують ціль дослідження і визначають набір економічних змінних моделі. При виборі економічних змінних необхідно здійснювати теоретичне обґрунтування кожної змінної. При цьому рекомендують, щоб їх число було не дуже великим і, як мінімум, у кілька разів меншим за число спостережень. Пояснюючі змінні не повинні бути пов'язані між собою функціональною або кореляційною залежністю, тому що це може призвести до неможливості оцінки параметрів моделі або до одержання нестійких оцінок, що не мають реального змісту, тобто до явища мультиколінеарності. Визначальним при включенні до моделі тих або інших змінних є економічний (якісний) аналіз досліджуваного об'єкта.

2-й етап (апріорний). На цьому етапі проводять аналіз сутності досліджуваного об'єкта, формування і формалізацію апріорної інформації.

3-й етап (параметризація). На цьому етапі здійснюють безпосередньо моделювання, тобто вибір загального вигляду моделі, виявлення зв'язків, які до неї входять. Основна задача, розв'язувана на цьому етапі, - вибір виду функції $f(x)$. Досить важливою проблемою на цьому етапі економетричного моделювання є проблема специфікації моделі, зокрема, вираження в математичній формі виявлених зв'язків і співвідношень; визначення складу **екзогенних** (незалежних) і **ендогенних** (залежних) змінних, формулювання вихідних передумов і обмежень моделі. Від того, наскільки вдало вирішена проблема специфікації моделі, у значній мірі залежить успіх всього економетричного моделювання.

4-й етап (інформаційний). На даному етапі здійснюють збір необхідної статистичної інформації - спостережуваних значень економічних змінних

$$(x_{i1}, x_{i2}, ..., x_{ip}; y_{i1}, y_{i2}, …, y_{iq}),\ i=1,\ n.$$

Тут можуть бути спостереження, отримані як за участю дослідника, так і без його участі (в умовах активного або пасивного експерименту).

5-й етап (ідентифікація моделі). Тут здійснюють статистичний аналіз моделі і оцінку її параметрів.

6-й етап (верифікація моделі). На цьому етапі проводять перевірку істинності, тобто адекватності моделі. З'ясовують, наскільки вдало вирішені проблеми специфікації, ідентифікації та ідентифікованості моделі, яка точність розрахунків, зроблених на підставі даної моделі, і, в остаточному підсумку, наскільки відповідає побудована модель реальному економічному об'єкту або процесу, який моделюють. Треба помітити, що коли є статистичні дані, які характеризують економічний об'єкт моделювання у поточний і попередній моменти часу, то для верифікації моделі, побудованої для прогнозу, досить порівняти реальні значення змінних в наступні моменти часу з відповідними їм значеннями, отриманими на основі розглянутої моделі за даними попередніх моментів.

Розглянутий поділ економетричного моделювання на окремі етапи носить певною мірою умовний характер, оскільки ці етапи можуть перетинатися, взаємно доповнювати один одного та ін.

13.3 Класифікація економетричних моделей

Нехай є p пояснюючих змінних (факторів) $x_1, x_2, ..., x_p$ і залежна змінна Y. Змінна Y є випадковою величиною, яка має при заданих значеннях факторів x_j певний розподіл. Зазвичай припускають, що умовні розподіли Y при кожному припустимому значенні факторів x_j - нормальні.

Пояснюючі змінні x_j ($j= 1,..., p$) можуть вважатися як випадковими, так і детермінованими, тобто приймаючими цілком певні значення.

Класична економетрична модель розглядає пояснюючі змінні x_j як детерміновані.

Пояснена частина Y_e в кожному разі є функцією від значень факторів - пояснюючих змінних:

$$Y_e=f(x_1, x_2, ..., x_p).$$

Отже, економетрична модель має вигляд

$$Y=f(x_1, x_2, ..., x_p)+\varepsilon. \qquad (13.1)$$

Найбільш природним вибором поясненої частини випадкової величини Y є її середнє значення - умовне математичне сподівання $M_x(Y)$, отримане при певному наборі значень пояснюючих змінних $(x_1, x_2, ..., x_p)$.

Рівняння $M_x(Y) = f(x_1, x_2, ..., x_p)$ називають **рівнянням регресії**.

При такому природному виборі поясненої частини економетрична модель має вигляд

$$Y= M_x(Y)+\varepsilon, \qquad (13.2)$$

де ε - випадкова величина, називана **збуренням** або **помилкою**.

В курсі математичної статистики рівняння (13.2) називають рівнянням регресійної моделі.

Відзначимо, що економетрична модель не обов'язково є регресійною, тобто пояснена частина не завжди є умовним математичним сподіванням залежної змінної. Однією з можливих причин того, що економетрична модель не регресійна, є наявність систематичних помилок виміру пояснюючих змінних. З математичної точки зору регресійні моделі опиняються істотно простішим об'єктом, чим економетрична модель загального типу (13.1).

Щоб одержати досить достовірні й інформативні дані щодо розподілу будь-якої випадкової величини, необхідно мати вибірку її спостережень досить великого обсягу. Така вибірка спостережень залежної змінної Y і пояснюючих змінних x_j $i=\overline{1,p}$ є відправною точкою будь-якого економетричного дослідження. Такі вибірки є наборами значень (x_{i1}, ..., x_{ip}; y_i), де $i=\overline{1,n}$; p - кількість пояснюючих змінних, n - число спостережень.

Як правило, число спостережень n досить велике й значно перевищує число p пояснюючих змінних. Проблема, однак, полягає в тому, що спостереження y_i, розглянуті в різних вибірках як випадкові величини Y_i і одержувані при різних наборах значень пояснюючих змінних x_j, мають в загальному випадку різний розподіл. Це означає, що для кожної випадкової величини Y_i ми маємо всього лише одно спостереження. Зрозуміло, на підставі одного спостереження не можна зробити адекватний висновок щодо розподілу випадкової величини, і потрібні додаткові припущення.

Прийнято виділяти три основних класи моделей, які застосовують для аналізу або прогнозу, це:

- регресійні моделі з одним рівнянням,
- моделі часових рядів,
- системи одночасних рівнянь.

Вибір моделі залежить від вигляду вихідних даних, на підставі яких будують економетричну модель, а також від характеру досліджуваного економічного процесу.

В класичному курсі економетрії розглядають два типи вибіркових даних: просторові дані й часові ряди.

Просторова вибірка. Прикладом просторових даних є, наприклад, набір відомостей (обсяг виробництва, кількість працівників, дохід та ін.) з різних фірм у той самий момент часу (просторовий зріз). Іншим прикладом можуть бути дані з курсів покупки/продажу наявної валюти в певний день в обмінних пунктах міста та ін.

В економіці під просторовою вибіркою розуміють набір показників економічних змінних, отриманий у певний момент часу. Для економетриста, однак, таке визначення не дуже зручне (через неоднозначність поняття «момент часу». Це може бути і день, і тиждень, і рік). Очевидно, про просторову вибірку має смисл говорити в тому випадку, якщо всі спостереження отримані приблизно в незмінних умовах, тобто являють собою набір незалежних вибіркових даних з певної генеральної сукупності.

Називатимемо просторовою вибіркою серію з n незалежних спостережень p-мірної випадкової величини (x_{i1},..., x_{ip}; y_i). При цьому надалі можна не розглядати X_j як випадкові величини. В цьому випадку різні випадкові величини Y_i виявляються між собою незалежними, що спричиняє некорельованість їх збурень, тобто коефіцієнт кореляції між збуреннями ε_i і ε_j

$$r(\varepsilon_i, \varepsilon_j) = 0 \quad \text{при} \quad i=j. \tag{13.3}$$

Умова (13.3) істотно спрощує модель і її статистичний аналіз.

На запитання, чи є вибірка серією незалежних спостережень, немає однозначної відповіді. Формально визначення незалежності випадкових величин, як правило, виявляється реально неперевіреним. Зазвичай за незалежні приймають величини, що не пов'язані причинно. Однак на практиці далеко не завжди питання про незалежність виявляється безперечним.

Економетрична модель, побудована на основі просторової вибірки експериментальних даних (x_i, y_i) є регресійною моделлю з одним рівнянням

$$y_i = f(x_i) + \varepsilon_i, \qquad i=1\ldots n, \tag{13.4}$$

де помилки регресії задовольняють умовам

$$M(\varepsilon_i) = 0, \tag{13.5}$$

$$r(\varepsilon_i, \varepsilon_j) = 0, \tag{13.6}$$

$$D(\varepsilon_i) = \sigma_i^2. \tag{13.7}$$

Відносно умови (13.7) можливі два випадки:

- $\sigma_i^2 = \sigma_j^2$ при всіх i та j. Властивість сталості дисперсій помилок регресії називають **гомоскедастичністю**. В цьому випадку розподіли випадкових величин Y_i відрізняються тільки значенням математичного сподівання (поясненої частини);

- $\sigma_i^2 \neq \sigma_j^2$. У цьому випадку має місце **гетероскедастичність** моделі. Гетероскедастичність «псує» багато результатів статистичного аналізу і, як правило, вимагає усунення.

В певних випадках гетероскедастичність моделі очевидна, однак у більшості випадків потрібне застосування методів математичної статистики для прийняття рішення про те, який тип моделі слід розглядати.

Залежно від вигляду функції $f(x_i)$ моделі поділяють на лінійні і нелінійні. Наприклад, можна досліджувати попит на морозиво як функцію від часу, від температури повітря, від середнього рівня доходів або залежність зарплати від віку, статі, рівня освіти, стажу роботи та ін. Область застосування таких моделей, заснованих на просторових даних, навіть лінійних, значно ширша, ніж моделей часових рядів.

Часовий (динамічний) ряд. Прикладами часових даних можуть бути щоквартальні дані з інфляції, середньої заробітної плати, національного доходу, грошової емісії за останні роки або, наприклад, щоденний курс долара США на МВБ, ціни ф'ючерсних контрактів на поставку долара США та ін.

Відмінною рисою часових даних є те, що вони природно впорядковані за часом, окрім того, спостереження в близькі моменти часу часто бувають залежними.

Часовим (динамічним) рядом називають вибірку спостережень, в якої важливі не тільки самі спостережувані значення випадкових величин, але й порядок їх проходження один за одним. Найчастіше впорядкованість зумовлена тим, що експериментальні дані є серією спостережень тієї самої випадкової величини в послідовні моменти часу. В цьому випадку динамічний ряд називають часовим рядом. При цьому покладають, що тип розподілу спостережуваної випадкової величини залишається тим самим (наприклад, нормальним), але параметри його змінюються залежно від часу.

Моделі часових рядів, як правило, складніші за моделі просторової вибірки, тому що спостереження у випадку часового ряду, загалом кажучи, не є незалежними, а це означає, що помилки регресії можуть корелювати одна з одною, тобто умова (13.3) не виконується. Це значно ускладнює статистичний аналіз моделі.

Відзначимо, що маючи тільки ряд спостережень без розуміння їх природи, неможливо визначити, маємо ми справу з просторовою вибіркою або з часовим рядом.

До класу моделей часових рядів належать моделі тренду, сезонності, а також тренду і сезонності.

Модель тренду має вигляд:

$$Y(t) = T(t) + \varepsilon_t, \qquad (13.8)$$

де $T(t)$ - часовий тренд заданого параметричного вигляду (наприклад, лінійний) $T(t) = a+bt$;

ε - випадкова (стохастична) компонента.

Модель сезонності має вигляд:

$$Y(t) = S(t) + \varepsilon_t, \qquad (13.9)$$

де $S(t)$ - періодична (сезонна) компонента.

Модель тренду і сезонності може бути адитивною або мультиплікативною моделлю:

$y(t) = T(t) + S(t) + \varepsilon_t$ - аддитивна,

$y(t) = T(t) * S(t) + \varepsilon_t$ - мультиплікативна,

де $T(t)$ - часовий тренд заданого параметричного виду;

$S(t)$ - періодична (сезонна) компонента.

До моделей часових рядів належить множина складніших моделей, таких, як моделі адаптивного прогнозу, моделі авторегресії і ковзного середнього та ін. Їх загальною рисою є те, що вони пояснюють поведінку часового ряду, виходячи тільки з його попередніх значень. Такі моделі можна застосовувати, наприклад, для вивчення й прогнозування обсягу продажів авіаквитків, попиту на морозиво, короткострокового прогнозу процентних ставок та ін.

Системи одночасних рівнянь. Моделі, які описують системами рівнянь, називають **системами одночасних рівнянь**. Системи можуть складатися з тотожностей і регресійних рівнянь, кожне з яких може, крім пояснюючих змінних, містити в собі пояснювальні змінні з інших рівнянь системи. Отже, тут ми маємо набір пояснювальних змінних, пов'язаних через рівняння системи. Класичним прикладом системи одночасних рівнянь є модель попиту Q^d і пропозиції Q^s. Коли попит на товар визначають його ціною P і доходом споживача I, пропозиція товару - його ціною P і досягається рівновага між попитом та пропозицією:

$$Q^s = \alpha_1 + \alpha_2 P + \varepsilon_1 \quad \text{(пропозиція)}, \tag{13.10}$$

$$Q^d = \beta_1 + \beta_2 P + \beta_3 I + \varepsilon_2 \quad \text{(попит)}, \tag{13.11}$$

$$Q^d = Q^s \quad \text{(рівновага)}. \tag{13.12}$$

В цій системі екзогенною (незалежною) змінною є дохід споживача I, а ендогенними (залежними) – попит (пропозиція) товару $Q^d = Q^s = Q$ і ціна товару (ціна рівноваги) P.

В іншій моделі попиту та пропозиції як пояснююча пропозицію змінна може бути не тільки ціна товару P в певний момент часу t, але й ціна товару в попередній момент часу t-1, тобто лагова ендогенна змінна:

$$Q^s_t = \alpha_1 + \alpha_2 P_t + \alpha_3 P_{t-1} + \varepsilon_1. \tag{13.13}$$

Системи одночасних рівнянь вимагають застосування складнішого математичного апарату. Вони можуть використовуватися для моделей странової економіки та ін.

Отже, економетрична модель дозволяє пояснити поведінку ендогенних змінних залежно від значень екзогенних і лагових ендогенних змінних, інакше кажучи, залежно від визначених змінних.

Зауважимо, що не всяка економіко-математична модель, що представляє математико-статистичний опис досліджуваного економічного об'єкта, може вважатися економетричною. Вона стає економетричною тільки в тому випадку, якщо відбиватиме цей об'єкт на основі емпіричних (статистичних) даних, що характеризують саме його.

Контрольні запитання

1. З якою метою проводять економетричні дослідження?
2. Поясніть складові загальної економетричної моделі.
3. Охарактеризуйте етапи економетричного моделювання.
4. Поясніть відмінність між ендогенними і екзогенними змінними.
5. У чому полягає ідентифікація і верифікація економетричної моделі?
6. Поясніть терміни «регресія», «умовне математичне сподівання» і «збурення».
7. Які класи моделей використовують для аналізу або прогнозу в економетрії?
8. Якими властивостями повинна володіти регресійна модель з одним рівнянням, побудована на основі просторової вибірки?
9. Поясніть терміни «гомоскедастичність» і «гетероскедастичність».
10. Яку вибірку спостережень називають часовим рядом?
11. Які економетричні моделі належать до систем одночасних рівнянь?

ТЕМА 14
МЕТОДИ ПОБУДОВИ ЗАГАЛЬНОЇ ЛІНІЙНОЇ МОДЕЛІ

14.1 Побудова загальної лінійної моделі

В практиці економічних досліджень наявні дані не завжди можна вважати вибіркою з багатомірної нормальної сукупності, коли одна з розглянутих змінних не є випадковою або коли лінія регресії явно не пряма та ін. В цих випадках намагаються визначити криву (або поверхню), що дає найкраще наближення до вихідних даних. Відповідні методи наближення одержали назву **регресійного аналізу**.

Методи і моделі регресійного аналізу займають центральне місце в математичному апараті економетрії. Задачами регресійного аналізу є встановлення форми залежності між змінними, оцінка функції регресії, оцінка невідомих значень (прогноз значень) залежної змінної.

В економіці в більшості випадків між змінними величинами існують залежності, в яких кожному значенню однієї змінної відповідає не якесь певне, а множина можливих значень іншої змінної. Інакше кажучи, кожному значенню однієї змінної відповідає певний (умовний) розподіл іншої змінної. Така залежність одержала назву **статистичної** (або стохастичної, імовірнісної).

Виникнення поняття статистичного зв'язку зумовлене тим, що залежна змінна піддана впливу ряду неконтрольованих або неврахованих факторів, а крім того тим, що вимір значень змінних неминуче супроводжується певними випадковими помилками. Прикладом статистичного зв'язку є залежність урожайності від кількості внесених добрив, продуктивності праці на підприємстві від його енергоозброєності та ін.

В силу неоднозначності статистичної залежності між Y і X становить інтерес усереднена за X залежність, тобто закономірність у зміні умовного математичного сподівання $M_x(Y)$ (математичного сподівання випадкової змінної Y, обчисленого в припущенні, що змінна X прийняла значення x) залежно від x.

Якщо залежність між двома змінними така, що кожному значенню однієї змінної відповідає певне умовне математичне сподівання (середнє значення) іншої, то вона називається **кореляційною**. Інакше кажучи, кореляційною залежністю між двома змінними називають функціональну залежність між значеннями однієї з них і умовним математичним сподіванням іншої.

Кореляційна залежність може бути подана у вигляді

$$M_x(Y) = \varphi(x). \quad (14.1)$$

В регресійному аналізі розглядають однобічну залежність випадкової змінної Y від однієї (або кількох) невипадкової незалежної змінної X. Така залежність може виникнути, наприклад, коли при кожному фіксованому значенні X відповідні значення Y піддані випадковому розкиду за рахунок дії ряду неконтрольованих факторів. Таку залежність Y від X називають **регресійною**. При цьому залежну змінну Y називають також функцією відгуку, пояснювальною, вихідною, результуючою, ендогенною змінною, результативною ознакою, а незалежну змінну X - пояснюючою, вхідною, предикторною, екзогенною змінною, фактором, регресором, факторною ознакою.

Рівняння (14.1) називається **рівнянням регресії**, функція $\varphi(x)$ - **функцією регресії**, а її графік - **лінією регресії**.

Для точного опису рівняння регресії необхідно знати умовний закон розподілу залежної змінної Y за умови, що змінна X прийме значення x ($X=x$). У статистичній практиці таку інформацію одержати, як правило, не вдається, тому що зазвичай дослідник має лише вибірку пар значень (x_i, y_i) обмеженого обсягу n. У цьому випадку мова може йти про оцінку (наближений вираз, апроксимацію) за вибіркою функції регресії. Такою оцінкою є вибіркова лінія регресії:

$$\hat{y} = \hat{\varphi}(x, b_0, b_1, ..., b_p), \tag{14.2}$$

де y - умовна (групова) середня змінної Y при фіксованому значенні змінної $X=x$;

b_0, b_1, …, b_p - параметри кривої.

Рівняння (14.2) називають вибірковим рівнянням регресії.

При правильно визначеній апроксимуючій функції $\varphi(x, b_0, b_1, …, b_p)$ із збільшенням обсягу вибірки ($n \to \infty$) вона сходитиметься за імовірністю до функції регресії $\varphi(x)$.

Модель парної регресії називають **простою** (з двома змінними) **економетричною моделлю**. Її застосовують, якщо завдання полягає у визначенні або в оцінці залежності пояснювальної змінної Y від одного фактору (пояснюючої змінної) X. При цьому вихідні дані є двома наборами значень (вектори) $x = (x_1, …, x_n)$ і $y = (y_1, …, y_n)$. Метою є одержання аналітичної залежності пояснювальної змінної Y від пояснюючої змінної x, яка щонайкраще відповідає вихідним даним.

Отже, парна регресія є регресією між двома змінними – y та x, тобто модель виду:

$$y = \hat{f}(x), \tag{14.3}$$

де y - залежна змінна (результативна ознака);

x - незалежна, або пояснююча, змінна (ознака-фактор).

Знак «^» означає, що між змінними x та y немає строгої функціональної залежності, тому практично в кожному окремому випадку величина y складається із двох доданків:

$$y = \hat{y}_x + \varepsilon, \tag{14.4}$$

де y – фактичне значення результативної ознаки;

$\hat{y}_x$ – теоретичне значення результативної ознаки, знайдене виходячи з рівняння регресії;

ε – випадкова величина, що характеризує відхилення реального значення результативної ознаки від теоретичного, знайденого за рівнянням регресії.

Випадкову величину ε називають також **збуренням**. Вона включає вплив не врахованих у моделі факторів, випадкових помилок і особливостей виміру. Її присутність в моделі породжено трьома джерелами: специфікацією моделі, вибірковим характером вихідних даних і особливостями виміру змінних.

В регресійному аналізі розглядають однобічну залежність випадкової змінної Y від однієї (або кількох) невипадкової незалежної змінної X.

Від правильно обраної специфікації моделі залежить величина випадкових помилок: вони тим менші, чим менше теоретичні значення результативної ознаки $\hat{y}_x$ відрізняються від фактичних даних y.

До помилок специфікації відносять неправильний вибір тієї або іншої математичної функції для $\hat{y}_x$ і не врахування у рівнянні регресії будь-якого істотного фактору, тобто використання парної регресії замість множинної.

Поряд з помилками специфікації можуть мати місце помилки вибірки, які зазвичай зумовлені неоднорідністю даних у вихідній статистичній сукупності. Це, як правило, має місце при вивченні економічних процесів. Якщо сукупність неоднорідна, то рівняння регресії не має практичного змісту. Для одержання доброго результату зазвичай виключають із сукупності одиниці з аномальними значеннями досліджуваних ознак.

Використання часової інформації також є вибіркою з усієї множини хронологічних дат. Змінивши часовий інтервал, можна одержати інші результати.

Найбільшою небезпекою в практичному використанні методів регресії є помилки виміру. Якщо помилки специфікації можна зменшити, змінюючи форму моделі (вигляд математичного виразу), а помилки вибірки - збільшуючи обсяг вихідних даних, то помилки виміру практично зводять нанівець усі зусилля з кількісної оцінки зв'язку між ознаками.

Особливо велику роль мають помилки виміру при дослідженні на макрорівні. Так, у дослідженнях попиту і споживання як пояснюючу змінну широко використовують «дохід населення». Разом з тим, статистичний вимір величини доходу пов’язаний з рядом труднощів і не позбавлений можливих помилок, наприклад, в результаті наявності схованих доходів.

Припускаючи, що помилки виміру зведені до мінімуму, основну увагу в економетричних дослідженнях приділяють помилкам специфікації моделі.

В парній регресії вибір вигляду математичної функції $\hat{y}_x = f(x)$ можна здійснити за трьома методами:

- графічним;
- аналітичним, тобто виходячи з теорії досліджуваного взаємозв'язку;
- експериментальним.

При вивченні залежності між двома ознаками графічний метод підбору виду рівняння регресії є досить наочним. Він заснований на побудові **поля кореляції**. Основні типи кривих, які використовують при кількісній оцінці зв'язків, зображені на рисунку 14.1.

Значну зацікавленість викликає аналітичний метод вибору типу рівняння регресії. Він заснований на вивченні матеріальної природи зв'язку досліджуваних ознак.

При обробці інформації на комп'ютері вибір виду рівняння регресії зазвичай здійснюють за експериментальним методом, тобто шляхом порівняння величини залишкової дисперсії $\sigma^2_{зал}$, яку розраховують при різних моделях.

Якщо рівняння регресії охоплює всі точки кореляційного поля, що можливо тільки при функціональному зв'язку, коли всі точки лежать на лінії регресії $\hat{y}_x = f(x)$, то фактичні значення результативної ознаки збігаються з теоретичними $y = \hat{y}_x$, тобто вони повністю зумовлені впливом фактора *x*. В цьому випадку залишкова дисперсія $\sigma^2_{зал} = 0$.

В практичних дослідженнях, як правило, має місце певне розсіювання точок кореляційного поля щодо лінії регресії. Воно зумовлене впливом інших факторів, які не враховані у рівнянні регресії. Іншими словами, мають місце відхилення фактичних даних від теоретичних $(y - \hat{y}_x)$. Величина цих відхилень лежить в основі розрахунку залишкової дисперсії:

$$\sigma^2_{çàë} = \frac{1}{n}\sum_{i=1}^{n}\left(y - \hat{y}_x\right)^2 . \qquad (14.5)$$

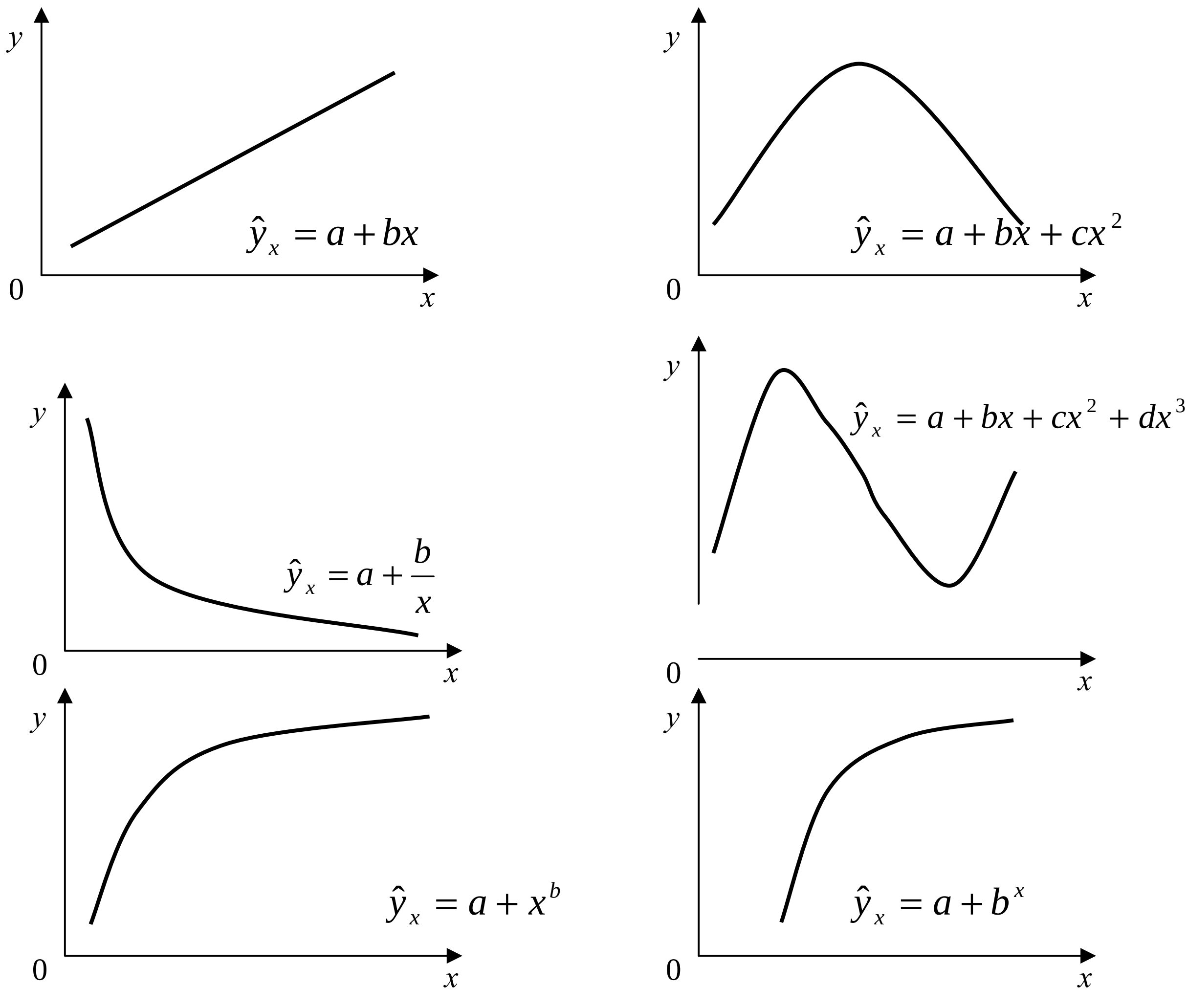


Рисунок 14.1 - Основні типи кривих, які використовують при кількісній оцінці зв'язків двох змінних

Чим меншою є величина залишкової дисперсії, тим меншим є вплив факторів, які не враховуються у рівнянні регресії, і тим краще рівняння регресії підходить до вихідних даних.

Вважають, що число спостережень має у 7-8 разів перевищувати число параметрів при змінній x, що розраховуються. Це означає, що шукати лінійну регресію, маючи менш за 7 спостережень, взагалі не має рації. Якщо вигляд функції ускладнюється, то необхідно збільшити обсяг спостережень, тому що кожний параметр при x повинен розраховуватися хоча б за 7 спостереженнями. Виходить, якщо ми вибираємо криву другого ступеня (параболу) $\hat{y}_x = a + bx + cx^2$, то потрібен обсяг інформації вже не менш за 14 спостережень.

14.2 Лінійна модель парної регресії. Суть методу найменших квадратів

Найкраще вивчені лінійні регресійні моделі, що задовольняють умовам (13.5), (13.6) і властивості сталості дисперсії помилок регресії, - їх називають **класичними моделями**.

Помітимо, що умовам класичної регресійної моделі задовольняють і гомоскедастична модель просторової вибірки, і модель часового ряду, спостереження якого не корелюють, а дисперсії постійні. З математичної точки зору вони дійсно нерозрізнювані (хоча можуть значно розрізнятися економічні інтерпретації отриманих математичних результатів). Нехай визначений характер експериментальних даних і виділений певний набір пояснюючих змінних. Для того щоб знайти пояснену частину, тобто величину $M_x(Y)$, потрібне знання умовних розподілів випадкової величини Y. На практиці це майже ніколи не має місця, тому точне знаходження поясненої частини неможливе. В таких випадках застосовують стандартну процедуру **згладжування експериментальних даних**. Ця процедура складається з двох етапів:

- визначають параметричне сімейство, до якого належить шукана функція $M_x(Y)$ (розглянута як функція від значень пояснюючих змінних X). Це може бути множина лінійних функцій, степеневих функцій та ін. (рис. 14.1);

- визначають оцінки параметрів цієї функції за допомогою одного з методів математичної статистики.

Формально ніяких способів вибору параметричного сімейства не існує. Однак у переважній більшості випадків економетричні моделі вибирають лінійними. Окрім цілком очевидної переваги лінійної моделі - її відносної простоти, - для такого вибору є, принаймні, дві істотні причини. Перша причина: якщо випадкова величина (X, Y) має спільний нормальний розподіл, тоді, як відомо, рівняння регресії лінійні. Припущення про нормальний розподіл є цілком природним і в ряді випадків може бути обґрунтованим за допомогою граничних теорем теорії імовірностей.

В інших випадках самі величини Y або X можуть не мати нормального розподілу, але певні функції від них розподілені нормально. Наприклад, відомо, що логарифм доходів населення - нормально розподілена випадкова величина.

Цілком природно, наприклад, вважати нормально розподіленою випадковою величиною пробіг автомобіля. Часто гіпотезу про нормальний розподіл приймають в багатьох випадках, коли немає явного їй протиріччя.

Друга причина, за якою лінійна регресійна модель опиняється переважніше інших, - це менший ризик значної помилки прогнозу. Лінійна регресія знаходить широке застосування в економетрії через чітку економічну інтерпретацію її параметрів.

Лінійна регресія зводиться до знаходження рівняння виду

$$\hat{y}_x = a + bx \text{ або } y = a + bx + \varepsilon . \qquad (14.6)$$

Рівняння виду $\hat{y}_x = a + bx$ дозволяє за заданим значенням фактору x знаходити теоретичні значення результативної ознаки, підставляючи до нього фактичні значення фактора x.

Побудова лінійної регресії зводиться до оцінки її параметрів – a і b. Класичний підхід до оцінювання параметрів лінійної регресії заснований на **методі найменших квадратів** (МНК), що відповідно до теореми Гауса-Маркова дає найкращі оцінки цих параметрів.

Теорема Гауса-Маркова. Якщо регресійна модель (14.6) має наступні властивості:

- збурення ε (або залежна змінна y_i) є величиною випадковою, а пояснююча змінна x_i - величиною невипадковою;
- математичне сподівання збурення ε дорівнює нулю;
- дисперсія збурення ε (або залежної змінної y_i) постійна для будь-якого i (або $D(y_i)=\sigma^2$) - умова гомоскедастичності або рівномінливості збурення (залежної змінної);
- збурення ε_i, і ε_j (або змінні y_i і y_j) не корельовані;
- збурення ε_i (або залежна змінна y_i) є нормально розподіленою випадковою величиною,

то оцінки параметрів лінійної регресії мають найменшу дисперсію в класі всіх лінійних незміщених оцінок.

Отже, оцінки параметрів лінійної регресії є **ефективними** оцінками.

МНК дозволяє одержати такі оцінки параметрів a і b, при яких сума квадратів відхилень фактичних значень результативної ознаки y від теоретичних $\hat{y}_x$ мінімальна:

$$\sum_{i=1}^{n} \left(y_i - \hat{y}_{xi}\right)^2 = \sum_{i=1}^{n} \varepsilon_i^2 \to \min . \qquad (14.7)$$

Тобто з усієї множини ліній лінію регресії на графіку вибирають так, щоб сума квадратів відстаней за вертикаллю між статистичними точками і цією лінією була би мінімальною (рис.14.2):

Як відомо з курсу математичного аналізу, щоб знайти мінімум функції (14.7), треба обчислити часткові похідні за кожним з параметрів a і b і дорівняти їх до нуля. Позначимо $\sum_{i=1}^{n} \varepsilon_i^2$ через $S(a, b)$, тоді:

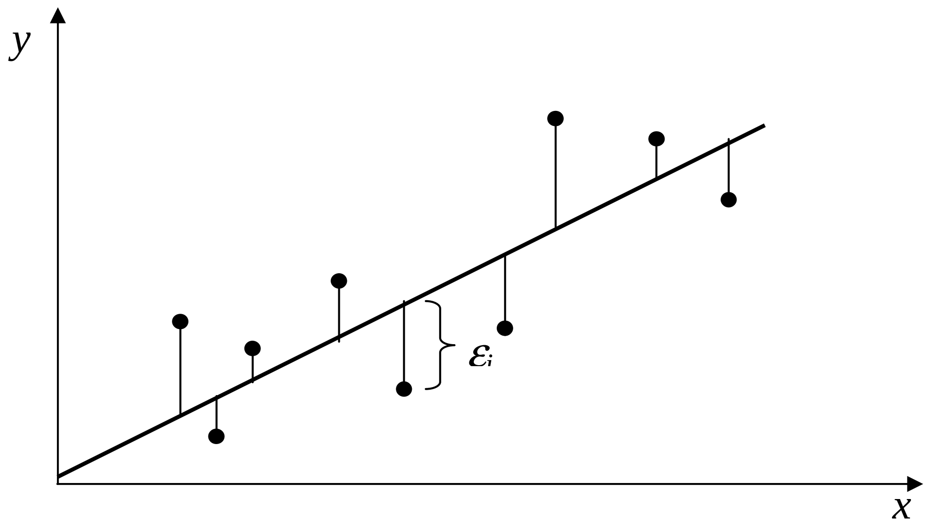


Рисунок 14.2 - Лінія регресії з мінімальною дисперсією залишків

$$S(a,b)=\sum_{i=1}^{n}(y-a-bx)^2,$$

$$\begin{cases} \dfrac{\partial S}{\partial a}=-2\sum\limits_{i=1}^{n}(y-a-bx)=0, \\ \dfrac{\partial S}{\partial b}=-2\sum\limits_{i=1}^{n}(y-a-bx)x=0. \end{cases} \tag{14.8}$$

Після нескладних перетворень одержимо наступну систему лінійних рівнянь для оцінки параметрів *a* і *b*:

$$\begin{cases} b*\sum x_i^2+a*\sum x_i=\sum x_i y_i, \\ b*\sum x_i+na=\sum y_i. \end{cases} \tag{14.9}$$

Вирішуючи систему рівнянь (14.9), знайдемо шукані оцінки параметрів *a* і *b*. Можна скористатися наступними готовими формулами, які випливають безпосередньо з розв'язання системи (14.9):

$$a=\overline{y}-b\overline{x}, \qquad b=\frac{\operatorname{cov}(x,y)}{\sigma_x^2}, \tag{14.10}$$

де $\operatorname{cov}(x,y)=\overline{yx}-\overline{y}*\overline{x}$ – коваріація ознак *x* і *y*;

$\sigma_x^2=\overline{x^2}-\overline{x}^2$ – дисперсія ознаки *x*

$$\overline{x}=\frac{1}{n}\sum_{i=1}^{n}x, \quad \overline{y}=\frac{1}{n}\sum_{i=1}^{n}y, \quad \overline{yx}=\frac{1}{n}\sum_{i=1}^{n}yx, \quad \overline{x^2}=\frac{1}{n}\sum_{i=1}^{n}x^2.$$

Коваріація - числова характеристика спільного розподілу двох випадкових величин, яка дорівнює математичному сподіванню добутку відхилень цих випадкових величин від їх математичних сподівань. Дисперсія - характеристика випадкової величини, що визначається як математичне сподівання квадрата відхилення випадкової величини від її математичного сподівання. Математичне сподівання - сума добутків значень випадкової величини та відповідних імовірностей.

Параметр *b* називають **коефіцієнтом регресії**. Його величина показує середню зміну результату зі зміною фактора на одну одиницю. Можливість чіткої економічної інтерпретації коефіцієнта регресії зробила лінійне рівняння регресії досить розповсюдженим в економетричних дослідженнях.

Формально *a* – значення *y* при *x*=0. Якщо фактор-ознака *x* не може мати нульового значення, то вищевказане трактування вільного члена *a* не має значення, тобто параметр *a* може не мати економічного змісту.

Рівняння регресії завжди доповнюється показником тісноти зв'язку результативної ознаки *Y* і фактора-ознаки *X*. При використанні лінійної регресії як такий показник виступає лінійний **коефіцієнт кореляції** r_{xy}, який можна розрахувати за наступними формулами:

$$r_{xy} = b\frac{\sigma_x}{\sigma_y} = \frac{\mathrm{cov}(x,y)}{\sigma_x \sigma_y}. \tag{14.11}$$

Лінійний коефіцієнт кореляції знаходиться в межах: $-1 \le r_{xy} \le 1$. Чим ближче абсолютне значення r_{xy} до одиниці, тим сильніше лінійний зв'язок між факторами (при $r_{xy} = \pm 1$ має місце строго функціональна залежність). Але необхідно мати на увазі, що близькість абсолютної величини лінійного коефіцієнта кореляції до нуля ще не означає відсутності зв'язку між ознаками. При іншій (нелінійній) специфікації моделі зв'язок між ознаками може виявитися досить тісним.

Для оцінки якості підбору лінійної функції розраховують квадрат лінійного коефіцієнта кореляції r_{xy}^2, який називають **коефіцієнтом детермінації**. Коефіцієнт детермінації характеризує частку дисперсії результативної ознаки *y*, що пояснюється регресією, в загальній дисперсії результативної ознаки:

$$r_{xy}^2 = 1 - \frac{\sigma_{зал}^2}{\sigma_y^2}, \tag{14.12}$$

де $\sigma_{\text{çàë}}^2 = \frac{1}{n}\sum_{i=1}^{n}(y - \hat{y}_x)^2$, $\sigma_y^2 = \frac{1}{n}\sum_{i=1}^{n}(y - \bar{y})^2 = \overline{y^2} - \bar{y}^2$.

Відповідно величина $1 - r_{xy}^2$ характеризує частку дисперсії *y*, викликану впливом інших, не врахованих в моделі, факторів.

Після того як рівняння лінійної регресії знайдене, проводять оцінку значущості як рівняння в цілому, так і окремих його параметрів.

14.3 Оцінка значущості рівняння лінійної регресії та перевірка моделі на адекватність за критеріями Стьюдента і Фішера

Перевірити значущість рівняння регресії - означає встановити, чи відповідає математична модель, що виражає залежність між змінними, експериментальним даним і чи достатньо включених до рівняння пояснюючих змінних (однієї або кількох) для опису залежної змінної.

Щоб мати загальне судження про якість моделі, визначають середню помилку апроксимації:

$$\overline{A} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y-\hat{y}_x}{y}\right| * 100\% . \qquad (14.13)$$

Середня помилка апроксимації має не перевищувати 8-10%.

Оцінка значущості рівняння регресії в цілому здійснюється за F-критерієм Фішера, розрахунку якого передує дисперсійний аналіз. В математичній статистиці дисперсійний аналіз розглядають як самостійний інструмент статистичного аналізу. В економетрії його застосовують як допоміжний засіб для вивчення якості регресійної моделі.

Відповідно до основної ідеї дисперсійного аналізу, загальну суму квадратів відхилень змінної y від середнього значення $\overline{y}$ розкладають на дві частини – пояснену і непояснену:

$$\sum_{i=1}^{n}(y-\overline{y})^2 = \sum_{i=1}^{n}(\hat{y}_x-\overline{y})^2 + \sum_{i=1}^{n}(y-\hat{y}_x)^2 , \qquad (14.14)$$

де $\sum_{i=1}^{n}(y-\overline{y})^2$ – загальна сума квадратів відхилень;

$\sum_{i=1}^{n}(\hat{y}_x-\overline{y})^2$ – сума квадратів відхилень, пояснена регресією (або факторна сума квадратів відхилень);

$\sum_{i=1}^{n}(y-\hat{y}_x)^2$ – залишкова сума квадратів відхилень, що характеризує вплив неврахованих у моделі факторів.

Схема дисперсійного аналізу має вигляд, наданий таблицею 14.1 (n - число спостережень, m - число параметрів при змінній x).

Таблиця 14.1 - Схема дисперсійного аналізу

Компоненти дисперсії	**Сума квадратів**	**Число ступенів свободи**	**Дисперсія на один ступінь свободи**
Загальна	$\sum_{i=1}^{n}(y-\overline{y})^2$	n-1	$S^2_{общ} = \frac{\sum_{i=1}^{n}(y-\overline{y})^2}{n-1}$
Факторна	$\sum_{i=1}^{n}(\hat{y}_x-\overline{y})^2$	m	$S^2_{факт} = \frac{\sum_{i=1}^{n}(\hat{y}_x-\overline{y})^2}{m}$
Залишкова	$\sum_{i=1}^{n}(y-\hat{y}_x)^2$	n-m-1	$S^2_{зал} = \frac{\sum_{i=1}^{n}(y-\hat{y}_x)^2}{n-m-1}$

Визначення дисперсії на один ступінь свободи призводить дисперсії до порівняного вигляду. Зіставляючи факторну і залишкову дисперсії, що розраховані на один ступінь свободи, одержимо величину F-критерію Фішера:

$$F = \frac{S^2_{факт}}{S^2_{зал}} . \qquad (14.15)$$

Фактичне значення F-критерію Фішера (14.15) порівнюють з табличним значенням $F_{табл}(\alpha,\ k_1,\ k_2)$ при рівні значущості α і ступенях свободи $k_1=m$ і $k_2=n-m-1$. При цьому якщо фактичне значення F-критерію перевищує табличне, то визнають статистичну значущість рівняння в цілому.

Для парної лінійної регресії m=1, тому

$$F=\frac{S^2_{факт}}{S^2_{зал}}=\frac{\sum(\hat{y}_x-\bar{y})^2}{\sum(y-\hat{y}_x)^2}*(n-2). \qquad (14.16)$$

Величина F-критерію пов'язана з коефіцієнтом детермінації r^2_{xy}, і її можна розрахувати за наступною формулою:

$$F=\frac{r^2_{xy}}{1-r^2_{xy}}*(n-2). \qquad (14.17)$$

В парній лінійній регресії оцінюють значущість не тільки рівняння в цілому, але й окремих його параметрів. З цією метою для кожного з параметрів визначають його стандартну помилку: m_a і m_b.

Стандартну помилку коефіцієнта регресії визначають за формулою:

$$m_b=\sqrt{\frac{S^2_{зал}}{\sum(x-\bar{x})^2}}=\frac{S_{зал}}{\sigma_x*\sqrt{n}}, \qquad (14.18)$$

де $S^2_{зал}=\frac{\sum(y-\hat{y}_x)^2}{n-2}$ – залишкова дисперсія на один ступінь свободи.

Величину стандартної помилки разом з t-розподілом Стьюдента при n-2 ступенях свободи застосовують для перевірки значущості коефіцієнта регресії і для розрахунку його довірчого інтервалу.

Для оцінки значущості коефіцієнта регресії його величину порівнюють з його стандартною помилкою, тобто визначають фактичне значення t-критерію Стьюдента: $t_b=\frac{b}{m_b}$, яке потім порівнюють з табличним значенням за певним рівнем значущості α і числом ступенів свободи n-2. Довірчий інтервал для коефіцієнта регресії визначають як $b \pm t_{табл} m_b$. Оскільки знак коефіцієнта регресії вказує на зростання результативної ознаки y при збільшенні ознаки-фактора x (b>0), зменшення результативної ознаки при збільшенні ознаки-фактора (b<0) або його незалежність від незалежної змінної (b=0), то межі довірчого інтервалу для коефіцієнта регресії не повинні містити суперечливих результатів (наприклад, $-1{,}5 \le b \le 0{,}8$). Такого роду запис вказує, що дійсне значення коефіцієнта регресії одночасно містить додатні і від'ємні величини і навіть нуль, чого не може бути.

Стандартну помилку параметра a визначають за формулою:

$$m_a=\sqrt{S^2_{зал}\frac{\sum x^2}{n\sum(x-\bar{x})^2}}=S_{зал}\frac{\sqrt{\sum x^2}}{\sigma_x n}. \qquad (14.19)$$

Процедура оцінювання значущості даного параметра не відрізняється від розглянутої вище для коефіцієнта регресії. Обчислюють t-критерій: $t_a = \frac{a}{m_a}$, його величину порівнюють з табличним значенням при n-2 ступенях свободи.

Значущість лінійного коефіцієнта кореляції перевіряють на основі величини помилки коефіцієнта кореляції m_r:

$$m_r = \sqrt{\frac{1-r^2}{n-2}}. \tag{14.20}$$

Фактичне значення t-критерію Стьюдента визначається як $t_r = \frac{r}{m_r}$.

Існує зв'язок між t-критерієм Стьюдента і F-критерієм Фішера:

$$t_b = t_r = \sqrt{F}. \tag{14.21}$$

В прогнозних розрахунках за рівнянням регресії визначають прогнозоване $\hat{y}_p$ значення як точковий прогноз $\hat{y}_x$ при $x_p = x_k$, тобто шляхом підстановки до рівняння регресії $\hat{y}_x = a + bx$ відповідного значення x. Він доповнюється розрахунком стандартної помилки $\hat{y}_p$, тобто $M(\hat{y}_p)$, і відповідною інтервальною оцінкою прогнозного значення $\hat{y}_p$:

$$\hat{y}_p - \Delta_{\hat{y}_p} \le \hat{y}_p \le \hat{y}_p + \Delta_{\hat{y}_p},$$

де $\Delta_{\hat{y}_p} = m_{\hat{y}_p} * t_{табл}$, а $m_{\hat{y}_p}$ – середня помилка прогнозованого індивідуального значення:

$$m_{\hat{y}_p} = S_{зал}\sqrt{1 + \frac{1}{n} + \frac{(x_p - \bar{x})^2}{n\sigma_x^2}}. \tag{14.22}$$

14.4 Лінійна модель множинної регресії

Економічні явища, як правило, визначаються великим числом одночасно і сукупно діючих факторів. У зв'язку із цим часто виникає задача дослідження залежності однієї залежної змінної Y від кількох пояснюючих змінних $X_1, X_2, ..., X_n$. Цю задачу вирішують за допомогою **множинного регресійного аналізу**. Парна регресія може дати добрий результат при моделюванні, якщо впливом інших факторів, можна зневажити. Якщо цим впливом зневажити не можна, то слід спробувати виявити вплив інших факторів, увівши їх до моделі, тобто побудувати рівняння множинної регресії

$$\hat{y} = f(x_1, x_2, ..., x_m), \tag{14.23}$$

де y – залежна змінна (результативна ознака);

x_i – незалежні, або пояснюючі, змінні (ознаки-фактори).

Множинну регресію широко використовують в розв’язанні проблем попиту, прибутковості акцій, при вивченні функції витрат виробництва, у макроекономічних розрахунках і цілому ряді інших питань економетрії. Основна мета множинної регресії - побудувати модель з великим числом факторів, визначивши при цьому вплив кожного з них окремо, а також сукупний їх вплив на показник, що моделюється.

У зв'язку з чіткою інтерпретацією параметрів найбільш широко використовують лінійну функцію. У лінійній множинній регресії $\hat{y}_x = a + b_1x_1 + b_2x_2 + ... + b_mx_m$ параметри при *x* називаються коефіцієнтами «чистої» регресії. Вони характеризують середню зміну результативної ознаки із зміною відповідного фактору на одиницю при незмінному значенні інших факторів, зафіксованих на середньому рівні.

Розглянемо лінійну модель множинної регресії

$$y = a + b_1x_1 + b_2x_2 + ... + b_mx_m + \varepsilon . \tag{14.24}$$

Класичний підхід до оцінювання параметрів лінійної моделі множинної регресії заснований на методі найменших квадратів (МНК). Теорема Гауса-Маркова, що розглянута для парної регресійної моделі, виявляється вірною і для моделі множинної регресії. Тобто МНК дає найбільш ефективні (тобто такі, що володіють найменшою дисперсією в класі лінійних незміщених оцінок) оцінки параметрів регресійної моделі.

Нагадаємо, що МНК дозволяє одержати такі оцінки параметрів, за якими сума квадратів відхилень фактичних значень результативної ознаки *y* від розрахункових $\hat{y}_x$ мінімальна:

$$\sum_i \left(y_i - \hat{y}_{x_i} \right)^2 \ \to \ \min \cdot \tag{14.25}$$

З виразу (14.25) дістанемо функцію *m*+1 аргументу:

$$S(a, b_1, b_2, ..., b_m) = \sum_{i=1}^{n} (y - a - b_1x_1 - b_2x_2 - ... - b_mx_m)^2 .$$

Визначивши часткові похідні першого порядку, дорівняємо їх до нуля:

$$\begin{cases} \dfrac{\partial S}{\partial a} = -2\sum_{i=1}^{n}(y - a - b_1x_1 - b_2x_2 - ... - b_mx_m) = 0, \\ \dfrac{\partial S}{\partial b_1} = -2\sum_{i=1}^{n}(y - a - b_1x_1 - b_2x_2 - ... - b_mx_m)x_1 = 0, \\ .... \\ \dfrac{\partial S}{\partial b_m} = -2\sum_{i=1}^{n}(y - a - b_1x_1 - b_2x_2 - ... - b_mx_m)x_m = 0. \end{cases}$$

Після елементарних перетворень приходимо до системи лінійних нормальних рівнянь для знаходження параметрів лінійного рівняння множинної регресії (14.24):

$$\begin{cases} na + b_1\sum_{i=1}^{n} x_{i1} + b_2\sum_{i=1}^{n} x_{i2} + \ldots + b_m\sum_{i=1}^{n} x_{im} = \sum_{i=1}^{n} y_i, \\ a\sum_{i=1}^{n} x_{i1} + b_1\sum_{i=1}^{n} x_{i1}^2 + b_2\sum_{i=1}^{n} x_{i1}x_{i2} + \ldots + b_m\sum_{i=1}^{n} x_{i1}x_{im} = \sum_{i=1}^{n} x_{i1}y_i, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ a\sum_{i=1}^{n} x_{im} + b_1\sum_{i=1}^{n} x_{i1}x_{im} + b_2\sum_{i=1}^{n} x_{i2}x_{im} + \ldots + b_m\sum_{i=1}^{n} x_{im}^2 = \sum_{i=1}^{n} x_{im}y_i. \end{cases} \quad (14.26)$$

Для двофакторної моделі дана система матиме вигляд:

$$\begin{cases} na + b_1\sum_{i=1}^{n} x_{i1} + b_2\sum_{i=1}^{n} x_{i2} = \sum_{i=1}^{n} y_i, \\ a\sum_{i=1}^{n} x_{i1} + b_1\sum_{i=1}^{n} x_{i1}^2 + b_2\sum_{i=1}^{n} x_{i1}x_{i2} = \sum_{i=1}^{n} x_{i1}y_i, \\ a\sum_{i=1}^{n} x_{i2} + b_1\sum_{i=1}^{n} x_{i1}x_{i2} + b_2\sum_{i=1}^{n} x_{i2}^2 = \sum_{i=1}^{n} x_{i2}y_i. \end{cases} \quad (14.27)$$

На практиці часто виникає необхідність у порівнянні впливу на результативну ознаку різних пояснюючих змінних. Для цього необхідно, щоб ці пояснюючі змінні були приведені до однакових одиниць виміру. В таких випадках використовують стандартизовані коефіцієнти регресії:

$$t_y = \frac{y - \bar{y}}{\sigma_y}. \quad (14.28)$$

Метод найменших квадратів можна застосувати і до рівняння множинної регресії в стандартизованому вигляді:

$$t_y = \beta_1 t_{x_1} + \beta_2 t_{x_2} + \ldots + \beta_m t_{x_m} + \varepsilon, \quad (14.29)$$

де $t_y, t_{x_1}, \ldots, t_{x_m}$ – стандартизовані змінні: $t_y = \frac{y - \bar{y}}{\sigma_y}$, $t_{x_i} = \frac{x_i - \bar{x}_i}{\sigma_{x_i}}$, для яких середнє значення дорівнює нулю: $\bar{t}_y = \bar{t}_{x_i} = 0$, а середнє квадратичне відхилення дорівнює одиниці: $\sigma_{t_y} = \sigma_{t_{xi}} = 1$;

β_i – стандартизовані коефіцієнти регресії.

Стандартизовані коефіцієнти регресії показують, на скільки одиниць зміниться в середньому результат, якщо відповідний фактор x_i зміниться на одну одиницю при незмінному середньому рівні інших факторів. В силу того, що всі змінні задані як центровані й нормовані, стандартизовані коефіцієнти регресії β_I можна порівнювати між собою. Порівнюючи їх один з одним, можна ранжирувати фактори за силою їх впливу на результативну ознаку. В цьому полягає основна перевага стандартизованих коефіцієнтів регресії на відміну від коефіцієнтів «чистої» регресії, які непорівнянні один з одним.

Застосовуючи МНК до рівняння множинної регресії в стандартизованому вигляді, одержимо систему нормальних рівнянь

$$\begin{cases} r_{yx_1} = \beta_1 + \beta_2 r_{x_1x_2} + \ldots + \beta_m r_{x_1x_m}, \\ r_{yx_2} = \beta_1 r_{x_1x_2} + \beta_2 + \ldots + \beta_m r_{x_1x_m}, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ r_{yx_m} = \beta_1 r_{x_1x_m} + \beta_2 r_{x_2x_m} + \ldots + \beta_m, \end{cases} \tag{14.30}$$

де r_{yx_i} і $r_{x_ix_i}$ – коефіцієнти парної і міжфакторної кореляції.

Коефіцієнти «чистої» регресії b_i зв'язані із стандартизованими коефіцієнтами регресії β_1 у такий спосіб:

$$b_i = \beta_i \frac{\sigma_y}{\sigma_{x_i}}. \tag{14.31}$$

Тому можна переходити від рівняння регресії в стандартизованому масштабі (14.29) до рівняння регресії в натуральному масштабі змінних (14.24), при цьому параметр a визначається як $a = \overline{y} - b_1\overline{x_1} - b_2\overline{x_2} - \ldots - b_m\overline{x}_m$.

Зміст стандартизованих коефіцієнтів регресії дозволяє використовувати їх при відсіванні факторів – з моделі виключають фактори з найменшим значенням β_1.

На основі лінійного рівняння множинної регресії (14.24) можуть бути знайдені часткові рівняння регресії:

$$\begin{cases} y_{x_1x_2,x_3,\ldots,x_m} = \hat{f}(x_1), \\ y_{x_2x_1,x_3,\ldots,x_m} = \hat{f}(x_2), \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ y_{x_mx_1,x_2,\ldots,x_{m-1}} = \hat{f}(x_m), \end{cases} \tag{14.32}$$

тобто рівняння регресії, які зв'язують результативну ознаку з відповідним фактором x_i при закріпленні інших факторів на середньому рівні. В розгорнутому вигляді систему (14.32) можна переписати так:

$$\begin{cases} y_{x_1x_2,x_3,\ldots,x_m} = a + b_1x_1 + b_2\overline{x}_2 + b_3\overline{x}_3 + \ldots + b_m\overline{x}_m + \varepsilon, \\ y_{x_2x_1,x_3,\ldots,x_m} = a + b_1\overline{x}_1 + b_2x_2 + b_3\overline{x}_3 + \ldots + b_m\overline{x}_m + \varepsilon, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ y_{x_mx_1,x_2,\ldots,x_{m-1}} = a + b_1\overline{x}_1 + b_2\overline{x}_2 + b_3\overline{x}_3 + \ldots + b_mx_m + \varepsilon. \end{cases}$$

При підстановці в ці рівняння середніх значень відповідних факторів вони приймають вигляд парних рівнянь лінійної регресії:

$$\begin{cases} y_{x_1 x_2, x_3, \dots, x_m} = A_1 + b_1 x_1, \\ y_{x_2 x_1, x_3, \dots, x_m} = A_2 + b_2 x_2, \\ \dots\dots\dots\dots\dots\dots\dots\dots \\ y_{x_m x_1, x_2, \dots, x_{m-1}} = A_m + b_m x_m, \end{cases} \tag{14.33}$$

де

$$\begin{cases} A_1 = a + b_2 \overline{x}_2 + b_3 \overline{x}_3 + \dots + b_m \overline{x}_m, \\ A_2 = a + b_1 \overline{x}_1 + b_3 \overline{x}_3 + \dots + b_m \overline{x}_m, \\ \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots \\ A_m = a + b_1 \overline{x}_1 + b_2 \overline{x}_2 + b_3 \overline{x}_3 + \dots + b_{m-1} x_{m-1}. \end{cases}$$

На відміну від парної регресії часткові рівняння регресії характеризують ізольований вплив фактора на результативну ознаку, оскільки інші фактори закріплені на незмінному рівні. Ефекти впливу інших факторів приєднані в них до вільного члена рівняння множинної регресії. Це дозволяє на основі часткових рівнянь регресії визначати часткові коефіцієнти еластичності:

$$\mathring{A}_{yx_i} = b_i \frac{x_i}{\hat{y}_{x_i x_1, x_2, \dots, x_{i-1}, x_{i+1}, \dots, x_m}}, \tag{14.34}$$

де b_i – коефіцієнт регресії для фактора x_i у рівнянні множинної регресії;

$\hat{y}_{x_i x_1, x_2, \dots, x_{i-1}, x_{i+1}, \dots, x_m}$ – часткове рівняння регресії.

Поряд з частковими коефіцієнтами еластичності можуть бути знайдені середні за сукупністю показники еластичності:

$$\overline{E}_i = b_i \frac{\overline{x}_i}{\overline{y}_{x_i}}, \tag{14.35}$$

які показують, на скільки відсотків у середньому зміниться результативна ознака при зміні відповідного фактора на 1%. Середні показники еластичності можна порівнювати один з одним і відповідно ранжирувати фактори за силою їх впливу на результативну ознаку.

14.5 Оцінка значущості множинної регресії і показники якості моделі

Практичну значущість рівняння множинної регресії оцінюють за допомогою **показника множинної кореляції** і його квадрата - **показника детермінації**.

Показник множинної кореляції характеризує тісноту зв'язку розглянутого набору факторів з досліджуваною ознакою або, інакше, оцінює тісноту загального впливу факторів на результативну ознаку.

Незалежно від форми зв'язку показник множинної кореляції може бути знайдений як індекс множинної кореляції:

$$R_{yx_1x_2...x_m} = \sqrt{1 - \frac{\sigma^2_{зал}}{\sigma^2_y}}\,, \tag{14.36}$$

де σ^2_y – загальна дисперсія результативної ознаки;

$\sigma^2_{зал}$ – залишкова дисперсія.

Межі зміни індексу множинної кореляції від 0 до 1. Чим ближче його значення до 1, тим тісніше зв'язок результативної ознаки з усім набором досліджуваних факторів. Величина індексу множинної кореляції повинна бути більше (або дорівнювати) максимального значення парного індексу кореляції:

$$R_{yx_1x_2...x_m} \geq r_{yx_i(\max)}\,, \qquad i = \overline{1, m}\,.$$

При правильному включенні факторів до регресійної моделі величина індексу множинної кореляції істотно відрізнятиметься від індексу кореляції парної залежності. Якщо додатково включені до рівняння множинної регресії фактори третьорядні, то індекс множинної кореляції може практично збігатися з індексом парної кореляції (розходження в третьому, четвертому знаках). Звідки випливає, що порівнюючи індекси множинної і парної кореляції, можна зробити висновок про доцільність включення до рівняння регресії того або іншого фактора.

Розрахунок індексу множинної кореляції передбачає визначення залишкової дисперсії:

$$\sigma^2_{\text{çàë}} = \frac{1}{n}\sum (y - \hat{y}_{x_1x_2...x_m})^2\,. \tag{14.37}$$

Можна користуватися формулою для індексу множинної детермінації:

$$R^2_{yx_1x_2...x_m} = 1 - \frac{\sum (y - \hat{y}_{x_1x_2...x_m})^2}{\sum (y - \overline{y})^2}\,. \tag{14.38}$$

При лінійної залежності ознак формула для індексу множинної кореляції може бути представленою в такий спосіб:

$$R_{yx_1x_2...x_m} = \sqrt{\sum_{i=1}^{n} \beta_i r_{yx_i}}\,, \tag{14.39}$$

де β_i – стандартизовані коефіцієнти регресії;

r_{yx_i} – парні коефіцієнти кореляції результативної ознаки з кожним з факторів.

Вираз індексу множинної кореляції для лінійної регресії називають **лінійним коефіцієнтом множинної кореляції**, або сукупним коефіцієнтом кореляції.

Сукупний коефіцієнт кореляції можна також визначити за допомогою матриці парних коефіцієнтів кореляції:

$$R_{yx_1x_2\ldots x_m} = \sqrt{1 - \frac{\Delta r}{\Delta r_{11}}}, \tag{14.40}$$

де

$$\Delta r = \begin{vmatrix} 1 & r_{yx_1} & r_{yx_2} & \ldots & r_{yx_p} \\ r_{yx_1} & 1 & r_{x_1x_2} & \ldots & r_{x_1x_p} \\ r_{yx_2} & r_{x_2x_1} & 1 & \ldots & r_{x_2x_p} \\ \ldots & \ldots & & \ldots & \ldots \\ r_{yx_p} & r_{x_px_1} & r_{x_px_2} & \ldots & 1 \end{vmatrix}$$

- визначник матриці парних коефіцієнтів кореляції;

$$\Delta r_{11} = \begin{vmatrix} 1 & r_{x_1x_2} & \ldots & r_{x_1x_p} \\ r_{x_2x_1} & 1 & \ldots & r_{x_2x_p} \\ \ldots & \ldots & \ldots & \ldots \\ r_{x_px_1} & r_{x_px_2} & \ldots & 1 \end{vmatrix}$$

- визначник матриці міжфакторної кореляції.

Як бачимо, величина множинного коефіцієнта кореляції залежить не тільки від кореляції результативної ознаки з кожним із факторів, але й від міжфакторної кореляції. Отже, визначити сукупний коефіцієнт кореляції можна, не звертаючись до рівняння множинної регресії, а використовуючи лише парні коефіцієнти кореляції.

У розглянутих показниках множинної кореляції (індекс і коефіцієнт) використовують значення залишкової дисперсії, що має систематичну помилку вбік зменшення. В результаті залишкова дисперсія тим менша, чим більше параметрів визначають у рівнянні регресії при заданому обсязі спостережень n. Якщо число параметрів при x_i дорівнює m і наближається до обсягу спостережень, то залишкова дисперсія наблизиться до нуля, і коефіцієнт (індекс) кореляції наблизиться до одиниці навіть при слабкому зв'язку факторів з результативною ознакою. Для усунення цього явища використовують скорегований індекс (коефіцієнт) множинної кореляції.

Скорегований індекс множинної кореляції містить виправлення на число ступенів свободи, а саме залишкова сума квадратів $\sum(y - \hat{y}_{x_1,x_2,\ldots,x_m})^2$ ділиться на число ступенів свободи залишкової варіації n-m-1, а загальна сума квадратів відхилень $\sum(y - \bar{y})^2$ - на число ступенів свободи в цілому за сукупністю (n-1).

Нагадаємо, що R^2 характеризує частку варіації залежної змінної, яка зумовлена регресією або мінливістю пояснюючих змінних. Чим ближче R^2 до одиниці, тим краще регресія описує залежність між пояснюючими змінними і результативною ознакою. Разом з тим використання тільки одного коефіцієнта детермінації R^2 для вибору найкращого рівняння регресії може виявитися недостатнім. На практиці зустрічаються випадки, коли погано визначена модель регресії може дати порівняно високий коефіцієнт R^2.

Недоліком коефіцієнта детермінації R^2 також є те, що він збільшується при додаванні нових пояснюючих змінних у модель, хоча це і не обов'язково поліпшує якість регресійної моделі. Тому переважніше використовувати скорегований коефіцієнт детермінації R^2, визначений за формулою

$$\hat{R}^2 = 1 - \frac{n-1}{n-p-1}(1-R^2)\,. \tag{14.41}$$

З (14.41) випливає, що чим більше кількість пояснюючих змінних p, тим менше $\hat{R}^2$ у порівнянні з R^2. На відміну від R^2 скорегований коефіцієнт $\hat{R}^2$ може зменшуватися при введенні до моделі нових пояснюючих змінних, які не роблять істотного впливу на залежну змінну. Однак навіть збільшення скорегованого коефіцієнта детермінації $\hat{R}^2$ при введенні до моделі нової пояснюючої змінної не завжди означає, що її коефіцієнт регресії є значущим (це відбувається тільки у випадку, якщо відповідне значення t-статистики більше за одиницю (за абсолютною величиною), тобто $|t|>1$. Інакше кажучи, збільшення $\hat{R}^2$ ще не означає поліпшення якості регресійної моделі.

Як було показано вище, ранжирування факторів, що беруть участь у множинній лінійній регресії, може бути проведене через стандартизовані коефіцієнти регресії (β-коефіцієнти). Ця сама мета може бути досягнута за допомогою часткових коефіцієнтів кореляції (для лінійних зв'язків). Крім того, часткові показники кореляції широко використовують при розв'язанні проблеми відбору факторів: доцільність включення того або іншого фактора до моделі можна довести величиною показника часткової кореляції.

Часткові коефіцієнти кореляції характеризують тісноту зв'язку між результативною ознакою і відповідним фактором при елімінуванні (усуненні впливу) інших факторів, включених до рівняння регресії.

Показники часткової кореляції являють собою відношення зменшення залишкової дисперсії за рахунок додаткового включення до аналізу нового фактора до залишкової дисперсії, що мала місце до введення його в модель.

В загальному вигляді при наявності m факторів для рівняння

$$y = a + b_1x_1 + b_2x_2 + \ldots + b_mx_m + \varepsilon$$

коефіцієнт часткової кореляції, що вимірює вплив на y фактору x_i, при незмінному рівні інших факторів, можна визначити за формулою:

$$r_{yx_1x_2\ldots x_{i-1}x_{i+1}\ldots x_m} = \sqrt{1 - \frac{1-R^2_{yx_1x_2\ldots x_i\ldots x_m}}{1-R^2_{yx_1x_2\ldots x_{i-1}x_{i+1}\ldots x_m}}}\,, \tag{14.42}$$

де $R^2_{yx_1x_2\ldots x_i\ldots x_m}$ – множинний коефіцієнт детермінації всіх m факторів з результативною ознакою;

$R^2_{yx_1x_2...x_{i-1}x_{i+1}...x_m}$ – той самий показник детермінації, але без введення до моделі фактору x_i.

При двох факторах формула (14.42) прийме вигляд:

$$r_{yx_1x_2} = \sqrt{1 - \frac{1 - R^2_{yx_1x_2}}{1 - R^2_{yx_2}}}, \qquad r_{yx_2x_1} = \sqrt{1 - \frac{1 - R^2_{yx_1x_2}}{1 - R^2_{yx_1}}}. \qquad (14.43)$$

Порядок часткового коефіцієнта кореляції визначають кількістю факторів, вплив яких виключається. Наприклад, $r_{yx_1x_2}$ – коефіцієнт часткової кореляції першого порядку. Відповідно коефіцієнти парної кореляції називають коефіцієнтами нульового порядку. Коефіцієнти часткової кореляції більш високих порядків можна визначити через коефіцієнти часткової кореляції більш низьких порядків за рекурентною формулою:

$$r_{yx_ix_1x_2...x_{i-1}x_{i+1}...x_m} = \frac{r_{yx_ix_1x_2...x_{i-1}x_{i+1}...x_{m-1}} - r_{yx_mx_1x_2...x_{m-1}} * r_{yx_ix_mx_1x_2...x_{i-1}x_{i+1}...x_{m-1}}}{\sqrt{\left(1 - r^2_{yx_mx_1x_2...x_{m-1}}\right) * \left(1 - r^2_{x_ix_mx_1x_2...x_{i-1}x_{i+1}...x_{m-1}}\right)}}. \qquad (14.44)$$

Для двох факторів дана формула прийме вигляд:

$$r_{yx_1x_2} = \frac{r_{yx_1} - r_{yx_2} * r_{x_1x_2}}{\sqrt{\left(1 - r^2_{yx_2}\right) * \left(1 - r^2_{x_1x_2}\right)}}, \qquad r_{yx_2x_1} = \frac{r_{yx_2} - r_{yx_1} * r_{x_1x_2}}{\sqrt{\left(1 - r^2_{yx_1}\right) * \left(1 - r^2_{x_1x_2}\right)}}. \qquad (14.45)$$

Для рівняння регресії з трьома факторами часткові коефіцієнти кореляції другого порядку визначають на основі часткових коефіцієнтів кореляції першого порядку. Так, за рівнянням $y = a + b_1x_1 + b_2x_2 + b_3x_3 + \varepsilon$ можливе вирахування трьох часткових коефіцієнтів кореляції другого порядку:

$$r_{yx_1x_2x_3}, \ r_{yx_2x_1x_3}, \ r_{yx_3x_1x_2},$$

кожний з яких визначають за рекурентною формулою. Наприклад, при *i=1* маємо формулу для розрахунку $r_{yx_1x_2x_3}$:

$$r_{yx_1x_2x_3} = \frac{r_{yx_1x_2} - r_{yx_3x_2} * r_{x_1x_2x_3}}{\sqrt{\left(1 - r^2_{yx_3x_2}\right) * \left(1 - r^2_{x_1x_3x_2}\right)}}. \qquad (14.46)$$

Розраховані за рекурентною формулою часткові коефіцієнти кореляції змінюються в межах від –1 до +1, а за формулами через множинні коефіцієнти детермінації – від 0 до 1. Порівняння їх одного з одним дозволяє ранжирувати фактори за тіснотою їх зв'язку з результативною ознакою. Часткові коефіцієнти кореляції дають міру тісноти зв'язку кожного фактора з результативною ознакою в чистому вигляді. Якщо із стандартизованого рівняння регресії $t_y = \beta_1 t_{x_1} + \beta_2 t_{x_2} + \beta_3 t_{x_3} + \varepsilon$ випливає, що $\beta_1 > \beta_2 > \beta_3$, тобто за силою впливу на результативну ознаку порядок факторів такий: x_1, x_2, x_3, то цей самий порядок факторів визначають і за співвідношенням часткових коефіцієнтів кореляції, $r_{yx_1x_2x_3} > r_{yx_2x_1x_3} > r_{yx_3x_1x_2}$.

В економетриці часткові коефіцієнти кореляції зазвичай не мають самостійного значення. Їх використовують на стадії формування моделі. Так, будуючи багатофакторну модель, на першому кроці визначають рівняння регресії з повним набором факторів і розраховують матрицю часткових коефіцієнтів кореляції. На другому кроці відбирають фактор з найменшою і несуттєвою за t-критерієм Стьюдента величиною показника часткової кореляції. Виключивши його з моделі, будують нове рівняння регресії. Процедура триває доти, поки не виявиться, що всі часткові коефіцієнти кореляції істотно відрізняються від нуля. Якщо виключено несуттєвий фактор, то множинні коефіцієнти детермінації на двох суміжних кроках побудови регресійної моделі майже не відрізняються один від одного, $R_{m+1}^2 \approx R_m^2$, де m – число факторів.

З наведених вище формул часткових коефіцієнтів кореляції видний зв'язок цих показників із сукупним коефіцієнтом кореляції. Знаючи часткові коефіцієнти кореляції (послідовно першого, другого і більш високого порядку), можна визначити сукупний коефіцієнт кореляції за формулою:

$$R_{yx_1x_2\ldots x_m} = \sqrt{1-\left(1-r_{yx_1}^2\right)*\left(1-r_{yx_2x_1}^2\right)*\left(1-r_{yx_3x_1x_2}^2\right)*\ldots*\left(1-r_{yx_mx_1x_2\ldots x_{m-1}}^2\right)}. \qquad (14.47)$$

Зокрема, для двофакторного рівняння формула (14.47) приймає вигляд:

$$R_{yx_1x_2\ldots x_m} = \sqrt{1-\left(1-r_{yx_1}^2\right)*\left(1-r_{yx_2x_1}^2\right)}. \qquad (14.48)$$

При повній залежності результативної ознаки від досліджуваних факторів коефіцієнт їх сукупного впливу дорівнює одиниці. Від одиниці віднімають частку залишкової варіації результативної ознаки ($1-r^2$), що зумовлена послідовно включеними до аналізу факторами. В результаті підкореневий вираз характеризує сукупну дію всіх досліджуваних факторів.

Значущість рівняння множинної регресії в цілому, так само як і в парній регресії, оцінюють за допомогою F-критерію Фішера:

$$F = \frac{S_{факт}}{S_{зал}} = \frac{R^2}{1-R^2} * \frac{n-m-1}{m}, \qquad (14.49)$$

де $S_{факт}$ – факторна сума квадратів на один ступінь свободи;

$S_{зал}$ – залишкова сума квадратів на один ступінь свободи;

R^2 – коефіцієнт (індекс) множинної детермінації;

m – число параметрів при змінних x (в лінійній регресії збігається з числом включених до моделі факторів);

n – число спостережень.

Оцінюють значущість не тільки рівняння в цілому, але й фактора, додатково включеного до регресійної моделі. Необхідність такої оцінки пов'язана з тим, що не кожен фактор, що ввійшов до моделі, може істотно збільшувати частку поясненої варіації результативної ознаки. Крім того, за наявності в моделі кількох факторів, їх можна вводити до моделі в різній послідовності. Через кореляцію між факторами значущість того самого фактора може бути різною залежно від послідовності його введення до моделі. Мірою для оцінки включення фактора до моделі служить частковий F-критерій, тобто F_{x_i}.

Частковий F-критерій побудований на порівнянні приросту факторної дисперсії, зумовленого впливом додатково включеного фактора, із залишковою дисперсією на один ступінь свободи за регресійною моделлю в цілому. В загальному вигляді для фактора x_i частковий F-критерій визначиться як

$$F_{x_i} = \frac{R^2_{yx_1...x_i...x_m} - R^2_{yx_1...x_{i-1}x_{i+1}...x_m}}{1 - R^2_{yx_1...x_i...x_m}} * \frac{n-m-1}{1}, \qquad (14.50)$$

де $R^2_{yx_1...x_i...x_m}$ – коефіцієнт множинної детермінації для моделі з повним набором факторів;

$R^2_{yx_1...x_{i-1}x_{i+1}...x_m}$ – той самий показник, але без включення до моделі фактора x_i;

n – число спостережень;

m – число параметрів у моделі (без вільного члена).

Фактичне значення часткового F-критерію порівнюють з табличним за рівнем значущості α і числом ступенів свободи: 1 і n-m-1. Якщо фактичне значення F_{x_i} перевищує $F(\alpha, k_1, k_2)$, то додаткове включення фактора x_i до моделі статистично виправдане і коефіцієнт чистої регресії b_i при факторі x_i статистично значущий. Якщо фактичне значення F_{x_i} менше табличного, то додаткове включення до моделі фактора x_i не збільшує істотно частку поясненої варіації ознаки y, отже, недоцільно його включення до моделі; коефіцієнт регресії при даному факторі в цьому випадку статистично не значущий.

Для двофакторного рівняння часткові F-критерії мають вигляд:

$$F_{x_1} = \frac{R^2_{yx_1x_2} - r^2_{yx_2}}{1 - R^2_{yx_1x_2}} * (n-3), \qquad F_{x_2} = \frac{R^2_{yx_1x_2} - r^2_{yx_1}}{1 - R^2_{yx_1x_2}} * (n-3). \qquad (14.51)$$

За допомогою часткового F-критерію можна перевірити значущість всіх коефіцієнтів регресії в припущенні, що кожний відповідний фактор x_i вводився до рівняння множинної регресії останнім.

Частковий F-критерій оцінює значущість коефіцієнтів чистої регресії. Знаючи величину F_{x_i}, можна визначити і t-критерій для коефіцієнта регресії при i-му факторі, t_{b_i}, а саме:

$$t_{b_i} = \sqrt{F_{x_i}}. \qquad (14.52)$$

Оцінка значущості коефіцієнтів чистої регресії за t-критерієм Стьюдента може бути проведена і без розрахунку часткових F-критеріїв. В цьому випадку, як і в парній регресії, для кожного фактора використовують формулу:

$$t_{b_i} = \frac{b_i}{m_{b_i}}, \qquad (14.53)$$

де b_i – коефіцієнт чистої регресії при факторі x_i;

m_{b_i} – середня квадратична (стандартна) помилка коефіцієнта регресії b_i.

Для рівняння множинної регресії $\hat{y} = a + b_1x_1 + b_2x_2 + ... + b_mx_m$ середню квадратичну помилку коефіцієнта регресії можна визначити за наступною формулою:

$$m_{b_1} = \frac{\sigma_y \sqrt{1 - R^2_{yx_1...x_m}}}{\sigma_{x_i} \sqrt{1 - R^2_{x_ix_1...x_m}}} \frac{1}{\sqrt{n - m - 1}}, \tag{14.54}$$

де σ_y – середнє квадратичне відхилення для ознаки y;

σ_{x_i} – середнє квадратичне відхилення для ознаки x_i;

$R^2_{yx_1...x_m}$ – коефіцієнт детермінації для рівняння множинної регресії;

$R^2_{x_ix_1...x_m}$ – коефіцієнт детермінації для залежності фактору x_i з усіма іншими факторами рівняння множинної регресії;

n-m-1 – число ступенів свободи для залишкової суми квадратів відхилень.

Як бачимо, щоб скористатися даною формулою, необхідно використати матрицю міжфакторної кореляції і розрахувати за нею відповідні коефіцієнти детермінації $R^2_{x_ix_1...x_m}$. Так, для рівняння $\hat{y} = a + b_1x_1 + b_2x_2 + b_3x_3$ оцінка значущості коефіцієнтів регресії b_1, b_2, b_3 припускає розрахунок трьох міжфакторних коефіцієнтів детермінації: $R^2_{x_1x_2x_3}$, $R^2_{x_2x_1x_3}$, $R^2_{x_3x_1x_2}$.

Взаємозв'язок показників часткового коефіцієнта кореляції, часткового F-критерію і t-критерію Стьюдента для коефіцієнтів чистої регресії може використовуватися в процедурі відбору факторів. Відсівання факторів при побудові рівняння регресії методом виключення практично можна здійснювати не тільки за частковими коефіцієнтами кореляції, виключаючи на кожному кроці фактор з найменшим незначущим значенням часткового коефіцієнта кореляції, але й за величинами t_{b_1} і F_{x_i}. Частковий F-критерій широко використовують і при побудові моделі за методом включення змінних і за кроковим регресійним методом.

Контрольні запитання

1. Поясніть терміни «рівняння регресії» і «функція регресії».
2. Які помилки належать до помилок специфікації? До помилок виміру?
3. На яких підставах вибирають вид регресійної залежності?
4. Як розраховують залишкову дисперсію? Загальну дисперсію ознаки y?
5. У чому полягає сутність методу найменших квадратів?
6. Якими властивостями повинна володіти лінійна модель, щоб оцінки її параметрів мали найменшу дисперсію в класі всіх лінійних незміщених оцінок?
7. Поясніть, що таке коефіцієнт регресії? Коефіцієнт кореляції? Коефіцієнт детермінації?

8. Як визначають статистичну значущість рівняння регресії в цілому?
9. З якою метою визначають стандартну помилку коефіцієнта регресії? Який критерій для цього використовують?
10. Як визначають довірчий інтервал для коефіцієнта регресії?
11. З якою метою і як визначають значущість лінійного коефіцієнта кореляції?
12. У якому випадку використовують лінійну модель множинної регресії?
13. Що таке коефіцієнти «чистої» регресії? Який в них економічний зміст?
14. У якому випадку використовують стандартизовані коефіцієнти регресії?
15. Що характеризують показник множинної кореляції і показник детермінації?
16. Що таке лінійний коефіцієнт множинної кореляції і скорегований індекс множинної кореляції?
17. Для чого розраховують часткові коефіцієнти кореляції?
18. Як оцінюють значущість рівняння множинної регресії в цілому та значущість коефіцієнтів чистої регресії?
19. Поясніть, у чому полягає процедура відбору факторів при побудові рівняння регресії методом виключення?

ТЕМА 15
МУЛЬТИКОЛІНЕАРНІСТЬ ТА ЇЇ ВПЛИВ НА ОЦІНКИ ПАРАМЕТРІВ МОДЕЛІ

15.1 Поняття мультиколінеарності. Вплив мультиколінеарності на оцінки параметрів

Однією з проблем, пов'язаних з практичним використанням моделі множинної регресії, є мультиколінеарність. Під **мультиколінеарністю** розуміють високу взаємну коррельованість пояснюючих змінних. Мультиколінеарність може проявлятися у **функціональній** (явній) і **стохастичній** (схованій) формах.

При функціональній формі мультиколінеарності принаймні один з парних зв'язків між пояснюючими змінними є лінійною функціональною залежністю. Це призводить до неможливості розв’язання відповідної системи нормальних рівнянь і одержання оцінок параметрів регресійної моделі.

Однак в економічних дослідженнях мультиколінеарність частіше проявляється в стохастичній формі, коли між хоча б двома пояснюючими змінними існує тісний кореляційний зв'язок. Оцінки параметрів стають дуже чутливими до незначної зміни результатів спостережень і обсягу вибірки. Рівняння регресії в цьому випадку, як правило, не мають реального змісту, тому що певні з коефіцієнтів можуть мати неправильні з погляду економічної теорії знаки і невиправдано великі значення.

Точних кількісних критеріїв для визначення наявності або відсутності мультиколінеарності не існує. Проте, є певні евристичні підходи до її виявлення. Один з таких підходів полягає в аналізі кореляційної матриці між пояснюючими змінними x_j і виявленні пар змінних, що мають високі коефіцієнти кореляції (зазвичай більші за 0,8). Якщо такі змінні існують, то говорять про мультиколінеарність між ними.

Корисно також знаходити множинні коефіцієнти детермінації між однією з пояснюючих змінних і певною групою з них. Наявність високого множинного коефіцієнта детермінації, що зазвичай перевищує 0,6, свідчить про мультиколінеарність.

Для усунення або зменшення мультиколінеарності використовують ряд методів. Найпростіший з них (але далеко не завжди можливий) полягає в тому, що із двох пояснюючих змінних, які мають високий коефіцієнт кореляції (більший за 0,8), одну змінну виключають з розгляду. При цьому яку змінну залишити, а яку видалити з аналізу, вирішують у першу чергу на підставі економічних міркувань. Якщо з економічної точки зору ні одній зі змінних не можна надати перевагу, то залишають ту із двох змінних, котра має більший коефіцієнт кореляції із залежною змінною. У цій вимозі проявляється специфіка множинної регресії як методу дослідження комплексного впливу факторів в умовах їх незалежності один від одного.

За величиною парних коефіцієнтів кореляції виявляється лише явна колінеарність факторів. Найбільші труднощі у використанні апарата множинної регресії виникають при наявності мультиколінеарності факторів, коли більш ніж два фактори зв'язані між собою лінійною залежністю, тобто має місце сукупний вплив факторів один на одного. Наявність мультиколінеарності факторів може означати, що певні фактори будуть завжди діяти в унісон. У результаті варіація у вихідних даних перестає бути повністю незалежною і не можна оцінити вплив кожного фактора окремо.

Включення до моделі мультиколінеарних факторів небажане в силу наступних наслідків:

- утруднюється інтерпретація параметрів множинної регресії як характеристик дії факторів в «чистому» вигляді, тому що фактори корельовані; параметри лінійної регресії втрачають економічний зміст;

- оцінки параметрів ненадійні, виявляють великі стандартні помилки й змінюються зі зміною обсягу спостережень (не тільки за величиною, але й за знаком), що робить модель непридатною для аналізу й прогнозування.

Для оцінки мультиколінеарності факторів може використовуватися визначник матриці парних коефіцієнтів кореляції між факторами.

Якщо фактори не корельовані між собою, то матриця парних коефіцієнтів кореляції між факторами є одиничною матрицею, оскільки всі недіагональні елементи $r_{x_i yj}$ ($i \neq j$) дорівнюють нулю. Так, для рівняння, що включає три пояснюючих змінних

$$\hat{y} = a + b_1 x_1 + b_2 x_2 + b_3 x_m$$

матриця коефіцієнтів кореляції між факторами має визначник, що дорівнює одиниці:

$$\mathrm{Det} = \begin{vmatrix} r_{x1x1} & r_{x1x2} & r_{x1x3} \\ r_{x2x1} & r_{x2x2} & r_{x2x3} \\ r_{x3x1} & r_{x3x2} & r_{x3x3} \end{vmatrix} = \begin{vmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{vmatrix} = 1.$$

Якщо між факторами існує повна лінійна залежність і всі коефіцієнти кореляції дорівнюють одиниці, то визначник такої матриці дорівнює нулю:

$$\mathrm{Det} = \begin{vmatrix} r_{x1x1} & r_{x1x2} & r_{x1x3} \\ r_{x2x1} & r_{x2x2} & r_{x2x3} \\ r_{x3x1} & r_{x3x2} & r_{x3x3} \end{vmatrix} = \begin{vmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{vmatrix} = 0.$$

Чим ближче до нуля визначник матриці міжфакторної кореляції, тим сильніша мультиколінеарність факторів і ненадійніші результати множинної регресії. І, навпаки, чим ближче до одиниці визначник матриці міжфакторної кореляції, тим менше мультиколінеарність факторів.

15.2 Методи виключення мультиколінеарності

Існує ряд підходів до подолання сильної міжфакторної кореляції. Найпростіший шлях усунення мультиколінеарності полягає у виключенні з моделі одного або кількох факторів. Інший підхід пов'язаний з перетворенням факторів, за яким зменшується кореляція між ними.

Одним із шляхів урахування внутрішньої кореляції факторів є перехід до сполучених рівнянь регресії, тобто до рівнянь, які відбивають не тільки вплив факторів, але і їх взаємодію. Так, якщо $\hat{y} = a + b_1x_1 + b_2x_2 + b_3x_m$, то можлива побудова сполученого рівняння наступного вигляду:

$$y = a + b_1x_1 + b_2x_2 + b_3x_3 + b_{12}x_1x_2 + b_{13}x_1x_3 + b_{23}x_2x_3 + \varepsilon .$$

Розглянуте рівняння включає взаємодію першого порядку (взаємодія двох факторів). Можливе включення до моделі і взаємодій більш високого порядку, якщо буде доведена їх статистична значущість за F-критерієм Фішера, але, як правило, взаємодії третього й більш високих порядків виявляються статистично незначущими.

Інший метод усунення або зменшення мультиколінеарності полягає в переході від незміщених оцінок, визначених за методом найменших квадратів, до зміщених оцінок, що володіють, однак, меншим розсіюванням оцінюваного параметра. Оцінки параметрів множинної регресії володіють відповідно до теореми Гауса-Маркова мінімальними дисперсіями в класі всіх лінійних незміщених оцінок, але при наявності мультиколінеарності ці дисперсії можуть виявитися занадто більшими, і звертання до відповідних зміщених оцінок може підвищити точність оцінювання параметрів регресії.

Для усунення мультиколінеарності може бути використаний перехід від вихідних пояснюючих змінних x_1, x_2,..., x_n, зв'язаних між собою досить тісною кореляційною залежністю, до нових змінних, які є лінійними комбінаціями вихідних. При цьому нові змінні повинні бути слабко корельованими або взагалі некорельованими. У якості таких змінних беруть, наприклад, так звані головні компоненти вектора вихідних пояснюючих змінних, досліджувані в компонентному аналізі, і розглядають регресію на головних компонентах, у якій останні виступають у якості узагальнених пояснюючих змінних, що підлягають надалі змістовній (економічній) інтерпретації.

Ортогональність головних компонентів запобігає прояві ефекту мультиколінеарності. Крім того, застосовуваний метод дозволяє обмежитися малим числом головних компонентів при порівняно великій кількості вихідних пояснюючих змінних.

Ще одним з можливих методів усунення або зменшення мультиколінеарності є використання **покрокових процедур відбору** найбільш інформативних змінних. Наприклад, на першому кроці розглядається лише одна пояснююча змінна, що має із залежною змінною y найбільший коефіцієнт детермінації. На другому кроці включається до регресії нова пояснююча змінна, котра разом з першою утворює пару пояснюючих змінних, що має з y найбільш високий (скорегований) коефіцієнт детермінації. На третьому кроці вводиться до регресії ще одна пояснююча змінна, котра разом із двома відібраними утворює трійку пояснюючих змінних, що має з y найбільший (скорегований) коефіцієнт детермінації, та ін.

Процедура введення нових змінних триває доти, поки збільшуватиметься відповідний (скорегований) коефіцієнт детермінації R^2 (точніше - мінімальне значення $R^2_{\min}$).

Розглянемо приклад. За даними n=20 сільськогосподарських районів області досліджується залежність змінної y - урожайності зернових культур (у ц/га) від ряду змінних - факторів сільськогосподарського виробництва: x_1 - число тракторів (приведеної потужності на 100 га); x_2 - число зернозбиральних комбайнів на 100 га; x_3 - число знарядь поверхневої обробки ґрунту на 100 га; x_4 - кількість добрив, що витрачаються на 1 га (т/га); x_5 - кількість хімічних засобів захисту рослин, що витрачаються на 1 га (ц/га). Вихідні дані наведені в таблиці 15.1.

Таблиця 15.1 – Вихідні дані до прикладу

i (номер району)	y_i	x_{i1}	x_{i2}	x_{i3}	x_{i4}	x_{i5}
1	9,70	1,59	0,26	2,05	0,32	0,14
2	8,40	0,34	0,28	0,46	0,59	0,66
19	13,10	0,08	0,25	0,03	0,73	0,20
20	8,70	1,36	0,26	0,17	0,99	0,42

У випадку виявлення мультиколінеарності вжити заходи з її усунення (зменшення), використовуючи покрокову процедуру відбору найбільш інформативних змінних.

З системи (14.26) знайдемо вектор оцінок параметрів регресійної моделі b= (3,515; -0,006; 15,542; 0,110; 4,475; -2,932), так що відповідно до (14.24) вибіркове рівняння множинної регресії має вигляд:

$$y = 33{,}515 - 0{,}006x_1 + 15{,}542x_2 + 0{,}110x_3 + 4{,}475x_4 - 2{,}932x_5,$$

де середні квадратичні відхилення (стандартні помилки) коефіцієнтів регресії b_i, обчислені за формулою (14.54), відповідно дорівнюють 5,41; 0,60; 21,59; 0,85; 1,54; 3,09.

Порівнюючи значення t-статистики (за абсолютною величиною) кожного коефіцієнта регресії b_i за формулою

$$t_{bj} = \frac{b_j}{\sigma_{bj}}, \; j = (0,1,2,3,4,5),$$

тобто t_{b0} = 0,65; t_{b1}= -0,01; t_{b2}= 0,72; t_{b3} = 0,13; t_{b4} = 2,91; t_{b5} =-0,95 із критичним значенням $t_{0,95;14}$= 2,14, визначеним з довідкової таблиці на рівні значущості α=0,05 при числі ступенів свободи $k = n-p-1 = 20 - 5 - 1 = 14$, бачимо, що значущим виявився тільки коефіцієнт регресії b_4 при змінній x_4 - кількість добрив, що витрачаються на гектар землі.

Множинний коефіцієнт детермінації врожайності зернових культур y за сукупністю п'яти факторів (x_1-x_5) сільськогосподарського виробництва виявився рівним R^2_{y12345} =0,517, тобто 51,7% варіації залежної змінної пояснюється включеними до моделі п'ятьма пояснюючими змінними. Оскільки обчислене фактичне значення F=3,00 більше табличного $F_{0,05;5;14}$=2,96, то рівняння регресії є значущим за F-критерієм на рівні α=0,05.

Розраховано матрицю парних коефіцієнтів кореляції, що наведені у таблиці 15.2.

Таблиця 15.2 - Парні коефіцієнти кореляції

Змінні	y	x_1	x_2	x_3	x_4	x_5
y	1,00	0,43	0,37	0,40	0,58*	0,33
x_1	0,43	1,00	0,85*	0,98*	0,11	0,34
x_2	0,37	0,85*	1,00	0,88*	0,03	0,46*
x_3	0,40	0,98*	0,88*	1,00	0,03	0,28
x_4	0,58*	0,11	0,03	0,03	1,00	0,57*
x_5	0,33	0,34	0,46*	0,28	0,57*	1,00

Знаком * позначені коефіцієнти кореляції, значущі за t-критерієм на 5%-му рівні.

Аналізуючи матрицю парних коефіцієнтів кореляції, можна відзначити тісний кореляційний зв'язок між змінними x_1 і x_2 (r_{12}= 0,85), x_1 і x_3 (r_{13} = 0,98), x_2 і x_3 (r_{23} = 0,88), що, свідчить про мультиколінеарність пояснюючих змінних.

Для усунення мультиколінеарності застосуємо процедуру **покрокового відбору** найбільш інформативних змінних.

1-й крок. З пояснюючих змінних x_1-x_5 виділяємо змінну x_4, що має із залежною змінною y найбільший коефіцієнт детермінації $R^2{}_{yj}$ (рівний для парної моделі квадрату коефіцієнта кореляції $r^2{}_{yj}$). Очевидно, це змінна x_4, тому що коефіцієнт детермінації $R^2{}_{y4}$ $=r_{y4}$ = $0{,}58^2$ =0,336 - максимальний. З урахуванням виправлення на незміщеність за формулою скорегований коефіцієнт детермінації $\hat{R}^2_{y4} = 1 - \frac{19}{18}(1 - 0{,}336) = 0{,}299$.

2-й крок. Серед усіляких пар пояснюючих змінних x_4, x_j, j =1,2,3,5, вибирається пара (x_4, x_3), що має із залежною змінною y найбільш високий коефіцієнт детермінації $R^2_{y4j} = R^2_{y43} = 0{,}483$ й з урахуванням виправлення $\hat{R}^2_{y43} = 1 - \frac{19}{17}(1 - 0{,}483) = 0{,}422$.

3-й крок. Серед усіляких трійок пояснюючих змінних (x_4, x_3, x_j), j = 1,2,5, найбільш інформативною виявилася трійка (x_4, x_3, x_5), що має максимальний коефіцієнт детермінації $R^2_{y43j} = R^2_{y435} = 0{,}513$ і відповідно скорегований коефіцієнт $\hat{R}^2_{y435} = 0{,}422$.

Оскільки скорегований коефіцієнт детермінації на 3-му кроці не збільшився, то в регресійній моделі досить обмежитися лише двома відібраними раніше пояснюючими змінними x_4 і x_3.

Рівняння регресії по цим змінним прийме вигляд:

$$\hat{y} = 7{,}29+3{,}48x_3+3{,}48x_4,$$

де середні квадратичні відхилення коефіцієнтів регресії b_j, обчислені за формулою (14.54), відповідно дорівнюють 0,66; 0,13; 1,07.

Неважко переконатися в тому, що тепер всі коефіцієнти регресії є значущими, тому що кожне зі значень t-статистики

$$t_{b0} = \frac{7{,}29}{0{,}66} = 11{,}0; \qquad t_{b3} = \frac{3{,}48}{0{,}13} = 26{,}8; \qquad t_{b4} = \frac{3{,}48}{1{,}07} = 3{,}25$$

перевищує відповідне табличне значення $t_{0,95;17}$=2,11.

Оскільки значення коефіцієнтів кореляції досить високі (понад 0,8): r_{12}=0,85, r_{13}=0,98, r_{23}=0,88, очевидно, що з відповідних трьох змінних x_1, x_2, x_3 дві змінні можна було відразу виключити з регресії й без проведення покрокового відбору, але як саме змінні виключити - треба вирішувати, виходячи з якісних міркувань, заснованих на знанні предметної області (у цьому випадку впливу на врожайність факторів сільськогосподарського виробництва).

Окрім розглянутої вище покрокової процедури приєднання пояснюючих змінних використовуються також покрокові процедури приєднання-видалення і процедура видалення пояснюючих змінних. Слід зазначити, що яка би покрокова процедура не використовувалася, вона не гарантує визначення оптимального (у смислі одержання максимального коефіцієнта детермінації R^2) набору пояснюючих змінних. Однак у більшості випадків одержувані за допомогою покрокових процедур набори змінних виявляються оптимальними або близькими до оптимальних.

Відбір факторів, що включаються до регресії, є одним з найважливіших етапів практичного використання методів регресії. Підходи до відбору факторів на основі показників кореляції можуть бути різними. Вони приводять побудову рівняння множинної регресії відповідно до різних методик. Залежно від того, яка методика побудови рівняння регресії прийнята, змінюється алгоритм її рішення на ЕОМ.

При відборі факторів рекомендують також користуватися наступним правилом: число факторів, що включають до моделі, зазвичай в 6-7 разів менше обсягу сукупності, за якою будують регресію. Якщо це співвідношення порушено, то число ступенів свободи залишкової дисперсії дуже мале. Це призводить до того, що параметри рівняння регресії виявляються статистично незначущими, а *F*-критерій менший за табличне значення.

Контрольні запитання

1. Поясніть поняття мультиколінеарності та у яких формах вона може проявлятися?
2. Які критерії використовують для визначення наявності або відсутності мультиколінеарності?
3. Охарактеризуйте методи, використовувані для усунення або зменшення мультиколінеарності.
4. Поясніть, у чому полягають наслідки включення до моделі мультиколінеарних факторів.
5. Як для оцінки мультиколінеарності факторів використовують визначник матриці парних коефіцієнтів кореляції між факторами?
6. Поясніть алгоритм покрокових процедур відбору найбільш інформативних змінних.

ТЕМА 16
УЗАГАЛЬНЕНИЙ МЕТОД НАЙМЕНШИХ КВАДРАТІВ

16.1 Поняття гомоскедастичності і гетероскедастичності

При оцінці параметрів рівняння регресії застосовують метод найменших квадратів (МНК). При цьому відповідно до теореми Гауса-Маркова виконують певні передумови щодо випадкової складової ε. У моделі

$$y = a + b_1x_1 + b_2x_2 + ... + b_mx_m + \varepsilon$$

випадкова складова ε є неспостережуваною величиною. Після того як зроблено оцінку параметрів моделі, розраховуючи різниці фактичних і теоретичних значень результативної ознаки y, можна визначити оцінки випадкової складової $y - \hat{y}_x$. Їх можна вважати певною вибірковою реалізацією невідомого залишку заданого рівняння, тобто ε_i.

При зміні специфікації моделі, додаванні до неї нових спостережень вибіркові оцінки залишків ε_i можуть змінюватися. Тому до задачі регресійного аналізу входить не тільки побудова самої моделі, але й дослідження випадкових відхилень ε_i, тобто залишкових величин.

При використанні критеріїв Фішера і Стьюдента роблять припущення щодо поведінки залишків ε_i, тобто що вони являють собою незалежні випадкові величини та їхнє математичне сподівання дорівнює 0, що вони мають однакову (постійну) дисперсію і підкоряються нормальному розподілу.

Статистична перевірка параметрів регресії і показників кореляції заснована на неперевіряємих передумовах розподілу випадкової складової ε_i. Вона носить лише попередній характер. Після побудови рівняння регресії проводять перевірку наявності у випадкових залишків ε_i передбачуваних властивостей. Це пов'язане з тим, що оцінки параметрів регресії повинні бути незміщеними, спроможними і ефективними. **Незміщеність** оцінки означає, що математичне сподівання залишків дорівнює нулю. Якщо оцінки параметрів мають властивість незміщеності, то їх можна порівнювати. Оцінки вважаються **ефективними**, якщо вони характеризуються найменшою дисперсією. У практичних дослідженнях це означає можливість переходу від точкового оцінювання параметрів до інтервального. **Спроможність** оцінок характеризує збільшення їх точності зі збільшенням обсягу вибірки. Великий практичний інтерес представляють ті результати регресійного аналізу, для яких довірчий інтервал очікуваного значення параметра регресії b_i має у границі значення імовірності, що дорівнює одиниці. Іншими словами, імовірність одержання оцінки на заданій відстані від істинного значення параметра близька до одиниці.

Метод найменших квадратів визначає оцінки регресії на основі мінімізації суми квадратів залишків. Тому дослідження залишків ε_i припускає перевірку наявності наступних п'яти передумов МНК:

- збурення ε є величина випадкова;

- математичне сподівання збурення ε дорівнює нулю;

- дисперсія збурення ε постійна для будь-якого i - умова гомоскедастичності або рівномінливості збурення;

- відсутність автокореляції залишків – значення залишків ε_i розподілені незалежно один від одного;

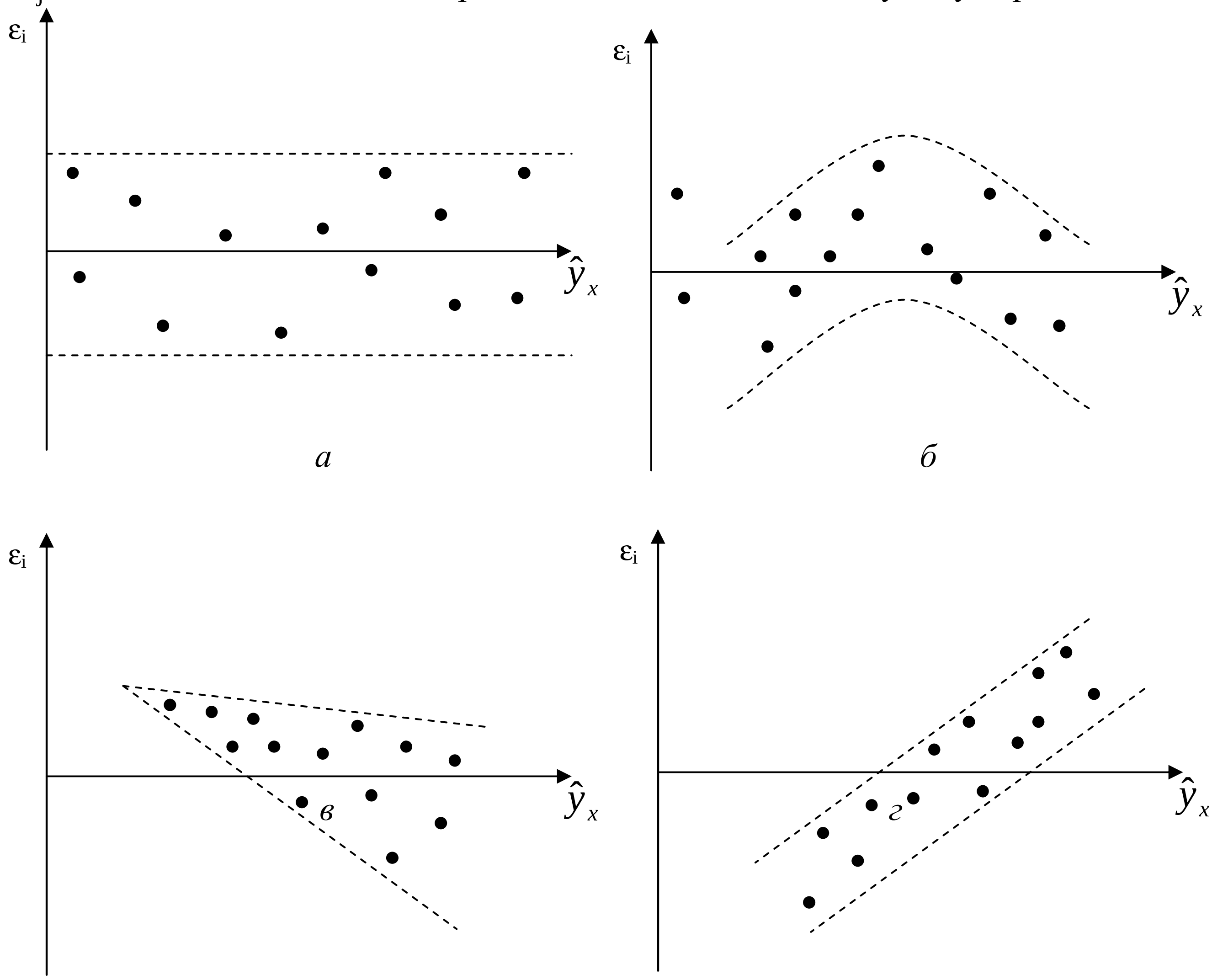
- збурення ε_i є нормально розподіленою випадковою величиною.

Якщо розподіл випадкових залишків ε_i не відповідає певним передумовам МНК, то необхідно корегувати модель. Насамперед, перевіряють випадковий характер залишків ε_i – перша передумова МНК. З цією метою будують графік залежності залишків ε_i від теоретичних значень результативної ознаки (рис. 16.1, *а*). Якщо графіком є горизонтальна смуга точок, то залишки ε_i являють собою випадкові величини і теоретичні значення $\hat{y}_x$ добре апроксимують фактичні значення y.

Друга передумова МНК щодо нульового значення математичного сподівання залишків означає, що $\sum(y - \hat{y}_x) = 0$. Це здійснимо для лінійних моделей і моделей, що нелінійні відносно включених змінних.

Разом з тим, незміщеність оцінок коефіцієнтів регресії, отриманих за МНК, залежить від незалежності випадкових залишків і величин x_j, що також досліджують в рамках дотримання другої передумови МНК. З цією метою поряд з розглянутим графіком залежності залишків ε_i від теоретичних значень результативної ознаки $\hat{y}_x$ будують графік залежності випадкових залишків ε_i від факторів, включених до регресії x_j (рис. 16.2).

Якщо залишки на графіку розташовані у вигляді горизонтальної смуги точок, то вони незалежні від значень x_j. Якщо графік показує наявність залежності ε_i і x_j, то модель неадекватна. Причини неадекватності можуть бути різними.



а - залишки ε_i є випадковими величинами; *б* - залишки ε_i не випадкові;

в - залишки ε_i не мають постійної дисперсії;

г - залишки ε_i мають систематичний характер

Рисунок 16.1 - Залежність випадкових залишків ε_i від теоретичних значень $\hat{y}_x$

Можливо, що порушено третю передумову МНК, і дисперсія залишків не постійна для кожного значення фактора x_j. Може бути невірною специфікація моделі, і до неї необхідно ввести додаткові члени від x_j, наприклад x_j^2. Скупчення точок у певних ділянках значень фактора x_j говорить про наявність систематичної помилки моделі.

Передумова про нормальний розподіл залишків дозволяє проводити перевірку параметрів регресії за допомогою F- і t-критеріїв. Разом з тим, оцінки параметрів регресії, знайдені із застосуванням МНК, мають гарні властивості навіть при відсутності нормального розподілу залишків, тобто при порушенні п'ятої передумови МНК.

Зовсім необхідним для одержання за МНК спроможних оцінок параметрів регресії є дотримання третьої і четвертої передумов.

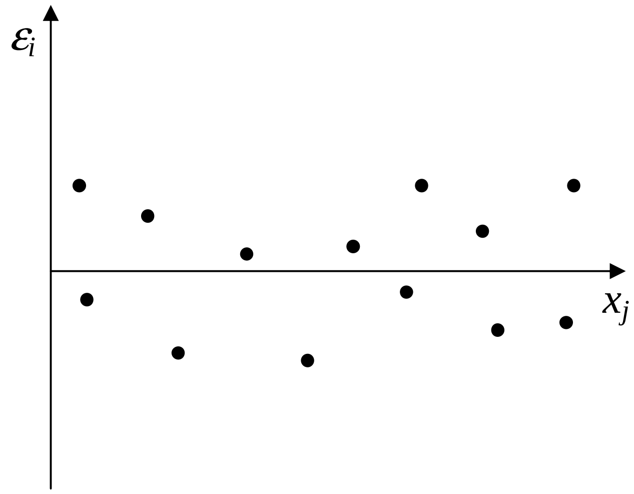


Рисунок 16.2 - Залежність величини залишків ε_i від величини фактору x_j

Відповідно до третьої передумови МНК потрібно, щоб дисперсія залишків була **гомоскедастичною**. Це значить, що для кожного значення фактору x_j залишки ε_i мають однакову дисперсію. Якщо ця умова застосування МНК не виконується, то має місце **гетероскедастичність**. Наявність гетероскедастичності показано на рисунку 16.3, *в*.

Для множинної регресії є найбільш прийнятний візуальний спосіб вивчення гомо- і гетероскедастичності на підставі графіків залежності залишків ε_i від теоретичних значень результативної ознаки $\hat{y}_x$.

При побудові регресійних моделей надзвичайно важливе дотримання четвертої передумови МНК, що полягає у відсутності **автокореляції** залишків, тобто значення залишків ε_i розподілені незалежно один від одного. Автокореляція залишків означає наявність кореляції між залишками поточних і попередніх (наступних) спостережень. Коефіцієнт кореляції між ε_i і ε_j, де ε_i – залишки поточних спостережень, ε_j – залишки попередніх спостережень (наприклад, $j=i-1$), може бути визначений як

$$r_{\varepsilon_i \varepsilon_j} = \frac{\operatorname{cov}(\varepsilon_i \varepsilon_j)}{\sigma_{\varepsilon_i} \sigma_{\varepsilon_j}},$$

тобто за звичайною формулою лінійного коефіцієнта кореляції. Якщо цей коефіцієнт виявиться істотно відмінним від нуля, то залишки автокорельовані і функція щільності імовірності $f(\varepsilon)$ залежить від j-ї точки спостереження і від розподілу значень залишків в інших точках спостереження.

Відсутність автокореляції залишкових величин забезпечує спроможність і ефективність оцінок коефіцієнтів регресії. Особливо актуальне дотримання даної передумови МНК при побудові регресійних моделей за рядами динаміки, де у зв'язку з наявністю тенденції наступні рівні динамічного ряду, як правило, залежать від своїх попередніх рівнів.

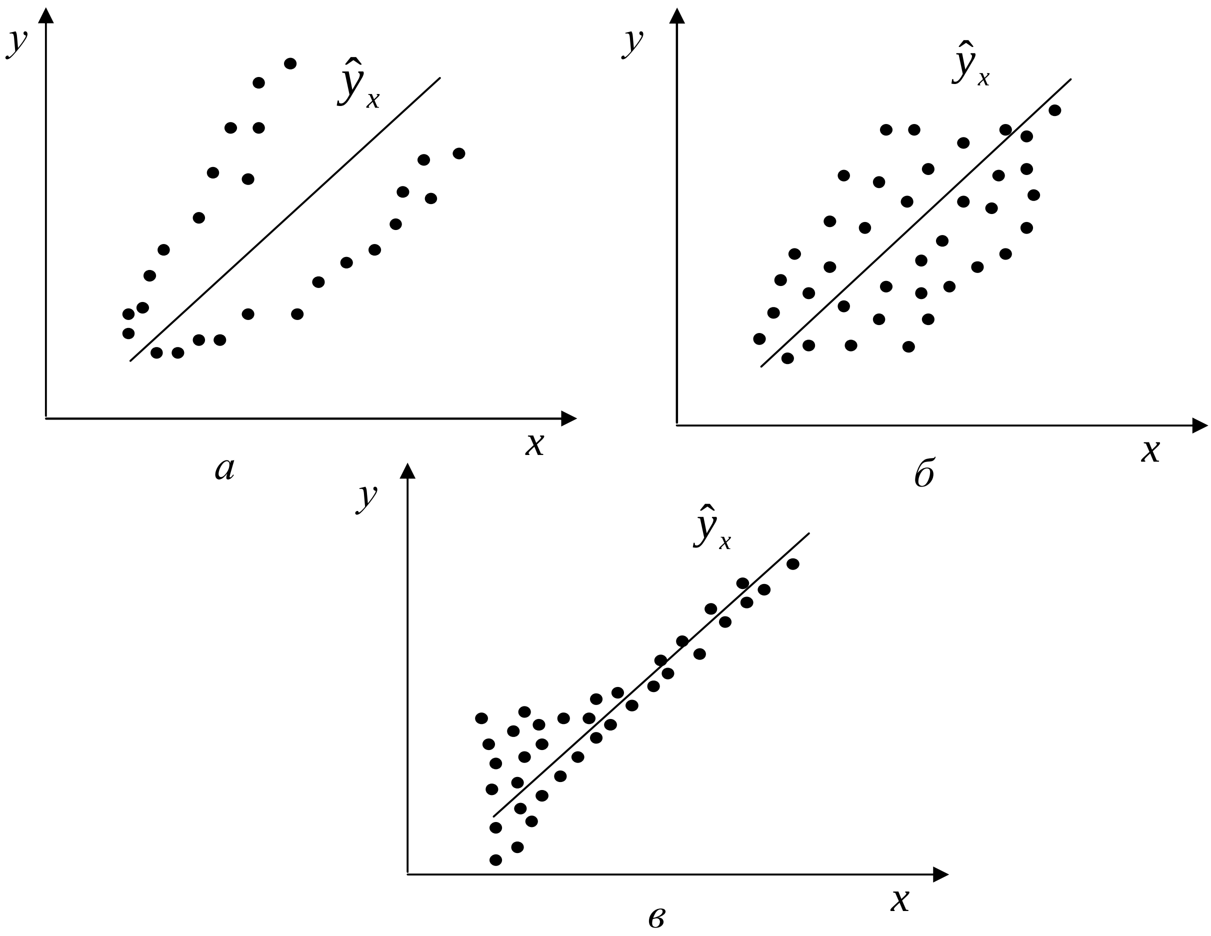


Рисунок 16.3 - Приклади гетероскедастичності

а - дисперсія залишків зростає в міру збільшення х; б - дисперсія залишків досягає максимальної величини при середніх значеннях змінної х і зменшується при мінімальних і максимальних значеннях х; в - дисперсія залишків є максимальною при малих значеннях х і стає однорідною в міру збільшення значень х

При недотриманні основних передумов МНК доводиться корегувати модель, змінюючи її специфікацію, додавати або виключати певні фактори, перетворювати вихідні дані для того, щоб одержати оцінки коефіцієнтів регресії, які мають властивість незміщеності, мають менше значення дисперсії залишків і забезпечують в зв'язку з цим більш ефективну статистичну перевірку значущості параметрів регресії.

16.2 Узагальнений метод найменших квадратів (УМНК)

При порушенні властивості гомоскедастичності та наявності автокореляції залишків ε_i рекомендують традиційний метод найменших квадратів заміняти узагальненим методом.

Узагальнений метод найменших квадратів застосовують до перетворених даних, він дозволяє одержувати оцінки, які володіють не тільки властивістю незміщеності, але й мають менші вибіркові дисперсії. Зупинимося на використанні УМНК для корегування гетероскедастичності.

Як і раніше, припустимо, що середнє значення залишкових величин дорівнює нулю. А от їх дисперсія не залишається незмінною для різних значень фактора, а пропорційна величині K_i, тобто

$$\sigma^2_{\varepsilon_i} = \sigma^2 K_i \,,$$

де $\sigma^2_{\varepsilon_i}$ – дисперсія помилки при конкретному і-му значенні фактору;

σ^2 – постійна дисперсія помилки при дотриманні передумови про гомоскедастичність залишків;

K_i – коефіцієнт пропорційності, що змінюється зі зміною величини фактора, що й зумовлює неоднорідність дисперсії.

При цьому вважають, що σ^2 невідома, а відносно величин K_i висувають певні гіпотези, що характеризують структуру гетероскедастичності.

У загальному виді для рівняння $y_i=a+bx_i+\varepsilon_i$ при $\sigma^2_{\varepsilon_i} = \sigma^2 K_i$ модель прийме вигляд: $y_i = a + bx_i + \sqrt{K_i \varepsilon_i}$. В ній залишкові величини є гетероскедастичними. Припускаючи в них відсутність автокореляції, можна перейти до рівняння з гомоскедастичними залишками, розділивши всі змінні, зафіксовані в ході і-го спостереження, на $\sqrt{K_i}$. Тоді дисперсія залишків буде величиною постійною, тобто $\sigma^2_{\varepsilon_i} = \sigma^2$.

Іншими словами, від регресії y на x ми перейдемо до регресії на нові змінні: $\frac{y}{\sqrt{K}}$ і $\frac{x}{\sqrt{K}}$. Рівняння регресії прийме вигляд:

$$\frac{y_i}{\sqrt{K_i}} = \frac{a}{\sqrt{K_i}} + b\frac{x_i}{\sqrt{K_i}} + \varepsilon_i \,,$$

а вихідні дані для цього рівняння матимуть вигляд:

$$y=\begin{vmatrix} \dfrac{y_1}{\sqrt{K_1}} \\ \dfrac{y_2}{\sqrt{K_2}} \\ \ldots\ldots \\ \dfrac{y_n}{\sqrt{K_n}} \end{vmatrix}, \qquad x=\begin{vmatrix} \dfrac{x_1}{\sqrt{K_1}} \\ \dfrac{x_2}{\sqrt{K_2}} \\ \ldots\ldots \\ \dfrac{x_n}{\sqrt{K_n}} \end{vmatrix}.$$

Щодо звичайної регресії рівняння з новими перетвореними змінними являє собою зважену регресію, в якій змінні y і x взяті з вагами $\frac{1}{\sqrt{K}}$.

Оцінка параметрів нового рівняння з перетвореними змінними приводить до зваженого методу найменших квадратів, для якого необхідно мінімізувати суму квадратів відхилень виду

$$S(a,b)=\sum_{i=1}^{n}\frac{1}{K_i}(y_i-a-bx_i)^2 .$$

Відповідно дістанемо наступну систему нормальних рівнянь:

$$\begin{cases} \sum\limits_{i=1}^{n}\dfrac{y_i}{K_i}=a\sum\limits_{i=1}^{n}\dfrac{1}{K_i}+b\sum\limits_{i=1}^{n}\dfrac{x_i}{K_i}, \\ \sum\limits_{i=1}^{n}\dfrac{y_i x_i}{K_i}=a\sum\limits_{i=1}^{n}\dfrac{x_i}{K_i}+b\sum\limits_{i=1}^{n}\dfrac{x_i^2}{K_i}. \end{cases}$$

Якщо перетворені змінні x і y взяти у відхиленнях від середніх рівнів, то коефіцієнт регресії b можна визначити як

$$b=\frac{\sum\limits_{i=1}^{n}\dfrac{1}{K_i}x_i y_i}{\sum\limits_{i=1}^{n}\dfrac{1}{K_i}x_i^2} .$$

При звичайному застосуванні методу найменших квадратів до рівняння лінійної регресії для змінних у відхиленнях від середніх рівнів коефіцієнт регресії b визначають за формулою:

$$b=\frac{\sum\limits_{i=1}^{n}x_i y_i}{\sum\limits_{i=1}^{n}x_i^2} .$$

Як бачимо, при використанні узагальненого МНК з метою корегування гетероскедастичності коефіцієнт регресії b являє собою зважену величину стосовно звичайного МНК з вагою $\frac{1}{\sqrt{K}}$.

Аналогічний підхід можливий не тільки для рівняння парної, але і для множинної регресії. Припустимо, що розглядають модель виду

$$y = a + b_1 x_1 + b_2 x_2 + \varepsilon\,,$$

для якої дисперсія залишкових величин виявилася пропорційною K_i^2. K_i являє собою коефіцієнт пропорційності, що приймає різні значення для відповідних i значень факторів x_1 і x_2. Через те, що $\sigma_{\varepsilon_i}^2 = \sigma^2 K_i^2$, розглянута модель прийме вигляд

$$y_i = a + b_1 x_{1i} + b_2 x_{2i} + K_i \varepsilon_i\,,$$

де залишки ε_i є гетероскедастичними.

Для того щоб одержати рівняння, де залишки ε_i є гомоскедастичними, слід перейти до нових перетворених змінних, розділивши всі члени вихідного рівняння на коефіцієнт пропорційності *K*. Рівняння з перетвореними змінними матиме вигляд

$$\frac{y_i}{K_i} = A + b_1 \frac{x_{1i}}{K_i} + b_2 \frac{x_{2i}}{K_i} + \varepsilon_i\,.$$

Це рівняння не містить вільного члена. Разом з тим, знайшовши змінні в новому перетвореному вигляді і застосовуючи звичайний МНК до них, одержимо іншу специфікацію моделі:

$$\frac{y_i}{K_i} = A + b_1 \frac{x_{1i}}{K_i} + b_2 \frac{x_{2i}}{K_i} + \varepsilon_i\,.$$

Параметри такої моделі залежать від концепції, прийнятої для коефіцієнта пропорційності K_i. В економетричних дослідженнях досить часто пропонують гіпотезу, що залишки ε_i пропорційні значенням фактора. Так, якщо у рівнянні

$$y = a + b_1 x_1 + b_2 x_2 + \ldots + b_m x_m + e$$

припустити, що $e = \varepsilon x_1$, тобто $K = x_1$ і $\sigma_{\varepsilon_i}^2 = \sigma^2 x_1$, то узагальнений МНК припускає оцінку параметрів наступного трансформованого рівняння:

$$\frac{y}{x_1} = b_1 + b_2 \frac{x_2}{x_1} + \ldots + b_m \frac{x_m}{x_1} + \varepsilon\,.$$

Застосування в цьому випадку узагальненого МНК приводить до того, що спостереження з меншими значеннями перетворених змінних $\frac{x}{K}$ мають при визначенні параметрів регресії відносно більшу вагу, ніж з первісними змінними. Разом з тим, треба мати на увазі, що нові перетворені змінні одержують новий економічний зміст і їх регресія має інший зміст, ніж регресія за вихідними даними.

Розглянемо приклад. Нехай *y* – витрати виробництва, x_1 – обсяг продукції, x_2 – основні виробничі фонди, x_3 – чисельність працівників, тоді рівняння

$$y = a + b_1 x_1 + b_2 x_2 + b_3 x_3 + e$$

є моделлю витрат виробництва з факторами обсягу. Припускаючи, що $\sigma_{\varepsilon_i}^2$ пропорційна квадрату чисельності працівників x_3, ми одержимо як результативну ознаку витрати на одного працівника $\frac{y}{x_3}$, а як фактори - наступні показники: продуктивність праці $\frac{x_1}{x_3}$ і фондоозброєність праці $\frac{x_2}{x_3}$. Відповідно трансформована модель матиме вигляд

$$\frac{y}{x_3} = b_3 + b_1\frac{x_1}{x_3} + b_2\frac{x_2}{x_3} + \varepsilon \ ,$$

де параметри b_1, b_2, b_3 чисельно не збігаються з аналогічними параметрами попередньої моделі. Крім цього, коефіцієнти регресії змінюють економічний зміст: з показників сили зв'язку, що характеризують середню абсолютну зміну витрат виробництва зі зміною абсолютної величини відповідного фактора на одиницю, вони фіксують при узагальненому МНК середню зміну витрат на працівника; зі зміною продуктивності праці на одиницю при незмінному рівні фондоозброєності праці; і зі зміною фондоозброєності праці на одиницю при незмінному рівні продуктивності праці.

Якщо припустити, що в моделі з первісними змінними дисперсія залишків пропорційна квадрату обсягу продукції, $\sigma^2_{\varepsilon_i} = \sigma^2 x^2_1$, можна перейти до рівняння регресії виду

$$\frac{y}{x_1} = b_1 + b_2\frac{x_2}{x_1} + b_3\frac{x_3}{x_1} + \varepsilon \ .$$

У ньому нові змінні: $\frac{y}{x_1}$ – витрати на одиницю (або на 1 грн. продукції), $\frac{x_2}{x_1}$ – фондомісткість продукції, $\frac{x_3}{x_1}$ – трудомісткість продукції.

Гіпотеза про те, що залишки пропорційні величині фактора, може мати реальну підставу: при обробці недостатньо однорідної сукупності, що включає як великі, так і малі підприємства, більшим значенням фактора обсягу можуть відповідати більші дисперсія результативної ознаки і дисперсія залишкових величин.

За наявності одної пояснюючої змінної гіпотеза $\sigma^2_{\varepsilon_i} = \sigma^2 x^2$ трансформує лінійне рівняння

$$y=a+bx+e$$

на рівняння

$$\frac{y}{x} = b + \frac{a}{x} + \varepsilon \ ,$$

у якому параметри a та b помінялися місцями, константа стала коефіцієнтом нахилу лінії регресії, а коефіцієнт регресії - вільним членом.

Перехід до відносних величин істотно знижує варіацію фактора і відповідно зменшує дисперсію помилки. Він являє собою найбільш простий випадок обліку гетероскедастичності в регресійних моделях за допомогою узагальненого МНК. Процес переходу до відносних величин може бути ускладнений висуванням інших гіпотез про пропорційність помилок щодо включених до моделі факторів. Використання тієї або іншої гіпотези припускає спеціальні дослідження залишкових величин для відповідних регресійних моделей. Застосування узагальненого МНК дозволяє одержати оцінки параметрів моделі, що володіють меншою дисперсією.

Контрольні запитання

1. Якими властивостями повинні володіти оцінки параметрів регресії?
2. Як впливає на параметри множинної лінійної моделі порушення умови, що математичне сподівання збурення ε дорівнює нулю?
3. Як впливає на параметри множинної лінійної моделі порушення умови, що дисперсія збурення ε постійною?
4. Який спосіб використовують для вивчення гомо- і гетероскедастичності?
5. На підставі якого дослідження роблять висновок про автокорельованість залишків?
6. Як поводяться при недотриманні основних передумов МНК?
7. Чим відрізняється узагальнений МНК від звичайного? У яких випадках його використовують?
8. Які властивості притаманні оцінкам параметрів моделі, отриманим на основі узагальненого МНК?

ТЕМА 17
ЕКОНОМЕТРИЧНІ МОДЕЛІ ДИНАМІКИ

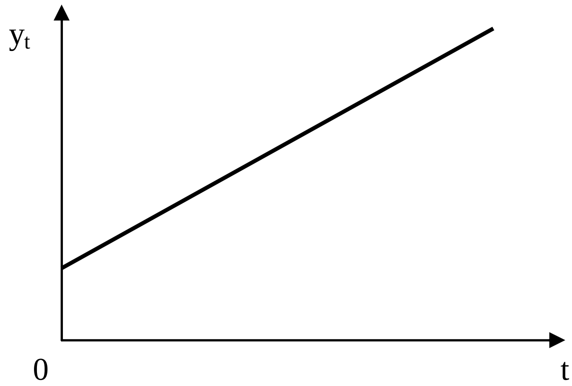
17.1 Загальні відомості про часові ряди і завдання їх аналізу

При розгляді класичної моделі регресії характер експериментальних даних, як правило, не має принципового значення. Однак це виявляється не так, коли порушені умови класичної моделі. Методи дослідження моделей, заснованих на даних просторових вибірок і часових рядів істотно відрізняються. Пояснюється це тим, що на відміну від просторових вибірок спостереження в часових рядах, як правило, не можна вважати незалежними.

Часовий ряд (ряд динаміки) – це сукупність значень будь-якого показника за кілька послідовних моментів або періодів часу. Кожний рівень часового ряду формується під впливом великої кількості факторів, які умовно можна підрозділити на три групи:

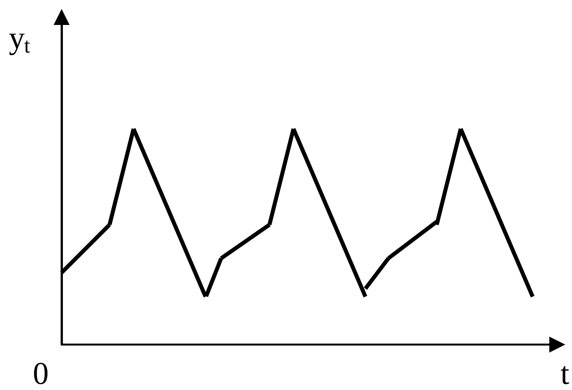
- фактори, що формують тенденцію ряду;
- фактори, що формують циклічні коливання ряду;
- випадкові фактори.

Більшість часових рядів економічних показників мають **тенденцію**, що характеризує сукупний довгостроковий вплив множини факторів на динаміку досліджуваного показника. Всі ці фактори, взяті окремо, можуть робити різнонаправлений вплив на досліджуваний показник. Однак у сукупності вони формують його зростаючу або спадну тенденцію. На рисунку 17.1 показаний гіпотетичний часовий ряд, що містить зростаючу тенденцію.



Рисунку 17.1 - Часовий ряд, що містить зростаючу тенденцію

Досліджуваний показник може також бути підданий циклічним коливанням. Ці коливання можуть носити сезонний характер, оскільки економічна діяльність ряду галузей економіки залежить від пори року. На рисунку 17.2 поданий гіпотетичний часовий ряд, що містить тільки сезонну компоненту.

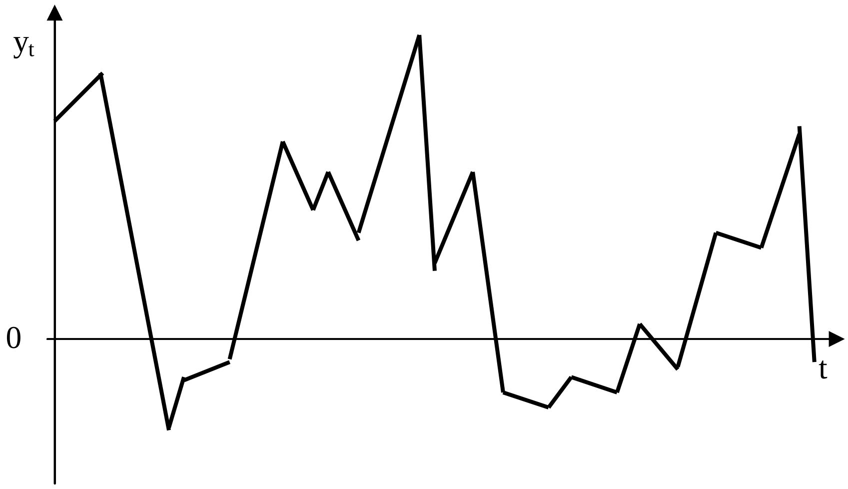


Рисунку 17.2 - Часовий ряд, що містить тільки сезонну компоненту

Певні часові ряди не містять тенденції і циклічної компоненти, а кожний наступний їх рівень утворюється як сума середнього рівня ряду і якоїсь (додатної або від’ємної) випадкової компоненти. Приклад ряду, що містить тільки випадкову компоненту, наведений на рисунку 17.3.

Очевидно, що реальні дані не випливають цілком з будь-яких описаних вище моделей. Найчастіше вони містять всі три компоненти. Кожний їх рівень формується під впливом тенденції, сезонних коливань і випадкової компоненти.

У більшості випадків фактичний рівень часового ряду можна представити як суму або добуток трендової, циклічної і випадкової компонентів. Модель, в якій часовий ряд представлений як сума перелічених компонентів, називається **адитивною моделлю** часового ряду. Модель, в якій часовий ряд представлений як добуток перелічених компонентів, називається **мультиплікативною моделлю** часового ряду. Основна задача економетричного дослідження окремого часового ряду - виявлення і побудова кількісного виразу кожної з перелічених вище компонентів для того, щоб використати отриману інформацію для прогнозування майбутніх значень ряду або при побудові моделей взаємозв'язку двох або більше часових рядів.



Рисунку 17.3 - Часовий ряд, що містить тільки випадкову компоненту

Найважливішою класичною задачею при дослідженні економічних часових рядів є виявлення і статистична оцінка основної тенденції розвитку досліджуваного процесу і відхилень від неї, або **тренду**.

Тренд або тенденція часового ряду - це певною мірою є умовним поняттям. Як тренд розуміють закономірну, невипадкову складову часового ряду (зазвичай монотонну), що може бути обчисленою за цілком визначеним однозначним правилом. Тренд часового ряду часто пов'язаний з дією фізичних законів або будь-яких інших об'єктивних закономірностей. Однак, загалом кажучи, не можна однозначно розділити випадковий процес або часовий ряд на регулярну частину (тренд) і коливальну частину (залишок). Тому зазвичай припускають, що тренд - це певна функція простого виду (лінійна, квадратична та ін.), що описує «поведінку в цілому» ряду або процесу. Якщо виділення такого тренда спрощує дослідження, то припущення про обрану форму тренда вважається припустимим.

Після виділення лінійного тренда потрібно з'ясувати, наскільки він значущий. Це робиться за допомогою аналізу коефіцієнту кореляції. Справа в тому, що відмінність коефіцієнта кореляції від нуля і тим самим наявність реального тренда (позитивного або негативного) може виявитися випадковою, зв'язаною із специфікою розглянутого відрізка часового ряду. Інакше кажучи, при аналізі іншого набору експериментальних даних (для того самого часового ряду) може виявитися, що отримана при цьому оцінка набагато ближче до нуля, ніж вихідна (і, можливо, навіть має інший знак), і говорити про реальний тренд тут уже стає важко. Такі ряди називаються часовими рядами з детерміністичним трендом. Існують також часові ряди із стохастичним трендом.

Відзначимо основні етапи аналізу часових рядів:

- графічне подання і опис поводження часового ряду;
- виділення і видалення закономірних (невипадкових) складових часового ряду (тренду, сезонних і циклічних складових);
- згладжування і фільтрація (видалення низько- або високочастотних складових часового ряду);

- дослідження випадкової складової часового ряду, побудова і перевірка адекватності математичної моделі для його опису;

- прогнозування розвитку досліджуваного процесу на основі наявного часового ряду;

- дослідження взаємозв'язку між різними часовими рядами.

Серед найпоширеніших методів аналізу часових рядів є моделі авторегресії і ковзної середньої, кореляційний і спектральний аналіз.

Якщо вибірку $y_1, y_2, ..., y_n$ розглядають як одну з реалізацій випадкової величини Y, часовий ряд $y_1, y_2, ..., y_n$ розглядається як одна з реалізацій (траєкторій) випадкового процесу $Y(t)$. Разом з тим необхідно мати на увазі принципові відмінності часового ряду y_t, $(t=1,n)$ від випадкової вибірки. По-перше, на відміну від елементів випадкової вибірки члени часового ряду, як правило, не є статистично незалежними. По-друге, члени часового ряду не є однаково розподіленими.

При наявності в часовому ряді тенденції і циклічних коливань значення кожного наступного рівня ряду залежать від попередніх. Кореляційну залежність між послідовними рівнями часового ряду називають **автокореляцією рівнів** ряду. Кількісно її можна виміряти за допомогою лінійного коефіцієнта кореляції між рівнями часового ряду, зрушеними на кілька кроків у часі. Формула для розрахунку коефіцієнта автокореляції має вигляд:

$$r_1 = \frac{\sum_{t=2}^{n}(y_t - \bar{y}_1)(y_{t-1} - \bar{y}_2)}{\sqrt{\sum_{t=2}^{n}(y_t - \bar{y}_1)^2 \sum_{t=2}^{n}(y_{t-1} - \bar{y}_2)^2}}, \qquad (17.1)$$

де $\bar{y}_1 = \frac{1}{n-1}\sum_{t=2}^{n} y_t$, $\bar{y}_2 = \frac{1}{n-1}\sum_{t=2}^{n} y_{t-1}$

Цю величину називають коефіцієнтом автокореляції рівнів ряду першого порядку, тому що він вимірює залежність між сусідніми рівнями ряду y_t і y_{t-1}.

Аналогічно можна визначити коефіцієнти автокореляції другого і більш високих порядків. Так, коефіцієнт автокореляції другого порядку характеризує тісноту зв'язку між рівнями y_t і y_{t-2} і визначають його за формулою:

$$r_1 = \frac{\sum_{t=3}^{n}(y_t - \bar{y}_3)(y_{t-2} - \bar{y}_4)}{\sqrt{\sum_{t=3}^{n}(y_t - \bar{y}_3)^2 \sum_{t=3}^{n}(y_{t-2} - \bar{y}_4)^2}}, \qquad (17.2)$$

де $\bar{y}_3 = \frac{1}{n-2}\sum_{t=3}^{n} y_t$, $\bar{y}_4 = \frac{1}{n-2}\sum_{t=3}^{n} y_{t-2}$.

Число періодів, за якими розраховується коефіцієнт автокореляції, називають ***лагом***. Із збільшенням лагу число пар значень, за якими розраховують коефіцієнт автокореляції, зменшується. Вважають за доцільне для забезпечення статистичної вірогідності коефіцієнтів автокореляції використовувати правило – максимальний лаг не повинен перевищувати $\tau \le \frac{n}{4}$.

Перелічимо властивості коефіцієнта автокореляції.

1. Його будують за аналогією з лінійним коефіцієнтом кореляції і в такий спосіб він характеризує тісноту тільки лінійного зв'язку поточного і попереднього рівнів ряду. Тому з коефіцієнта автокореляції можна судити про наявність лінійної (або близької до лінійної) тенденції. Для певних часових рядів, що мають сильну нелінійну тенденцію (наприклад, параболу другого порядку або експоненту), коефіцієнт автокореляції рівнів вихідного ряду може наближатися до нуля.

2. За знаком коефіцієнта автокореляції не можна робити висновок про зростаючу або спадну тенденцію в рівнях ряду. Більшість часових рядів економічних даних містять додатну автокореляцію рівнів, однак при цьому можуть мати спадну тенденцію.

Послідовність коефіцієнтів автокореляції рівнів першого, другого і т. д. порядків називають **автокореляційною функцією** часового ряду. Графік залежності її значень від величини лагу (порядку коефіцієнта автокореляції) називають **корелограмою.**

Аналіз автокореляційної функції і корелограми дозволяє визначити лаг, при якому автокореляція є найбільш високою, а отже, і лаг, при якому зв'язок між поточним і попереднім рівнями ряду найбільш тісний, тобто за допомогою аналізу автокореляційної функції і корелограми можна виявити структуру ряду.

Якщо найбільш високим виявився коефіцієнт автокореляції першого порядку, досліджуваний ряд містить тільки тенденцію. Якщо найбільш високим виявився коефіцієнт автокореляції порядку τ, то ряд містить циклічні коливання з періодичністю τ моментів часу. Якщо жоден з коефіцієнтів автокореляції не є значущим, можна зробити одне з двох припущень щодо структури цього ряду: або ряд не містить тенденції і циклічних коливань, або ряд містить сильну нелінійну тенденцію, для виявлення якої потрібно провести додатковий аналіз. Тому коефіцієнт автокореляції рівнів і автокореляційну функцію доцільно використовувати для виявлення в часовому ряді наявності або відсутності трендової компоненти і циклічної (сезонної) компоненти.

17.2 Моделювання тенденції часового ряду

Розповсюдженим способом моделювання тенденції часового ряду є побудова аналітичної функції, що характеризує залежність рівнів ряду від часу, або тренду. Цей спосіб називають аналітичним вирівнюванням часового ряду.

Оскільки залежність від часу може приймати різні форми, для її формалізації можна використати різні види функцій. Для побудови трендів найчастіше застосовують наступні функції, що наведені у таблиці 17.1.

Таблиця 17.1 - Види функцій

Вид рівняння	**Відбивана рівнянням тенденція**
Рівняння прямої $\overline{y}_t = at + b$	Рівномірне зростання при $a > 0$ або рівномірне падіння при $a < 0$
Експонентна функція $\overline{y}_t = a^t b$	Прискорюване зростання при $a > 1$ або падіння, що вповільнюється, при $a < 1$
Гіпербола $\overline{y}_t = \frac{a}{t} + b$	Вповільнюване падіння, при $a > 0$ або вповільнюване зростання, при $a < 0$
Парабола $\overline{y}_t = at^2 + bt + c$	Зростання, що переходить у падіння, або падіння, що переходить у зростання у точці $t = \frac{-b}{2a}$

Параметри наведених трендів можна визначити за звичайним методом найменших квадратів (МНК), використовуючи як незалежну змінну час t=1,n, а як залежну змінну – фактичні рівні часового ряду $\hat{y}_t$.

$$y_t = f(t) + \varepsilon_i, \tag{17.3}$$

де ε_i – збурення, що задовольняють основним передумовам регресійного аналізу, тобто уявляють собою незалежні і однаково розподілені випадкові величини, розподіл яких передбачається нормальним.

Нагадаємо, що відповідно до МНК параметри лінійного тренду $\hat{y}_t = f(t) = b_0 + b_1 t$ знаходять з системи нормальних рівнянь, в якій у якості пояснюючої змінної x_i фігурує час t.

$$\begin{cases} b_0 n + b_1 \sum_{t=1}^{n} t = \sum_{i=1}^{n} y_i \\ b_0 \sum_{t=1}^{n} t + b_1 \sum_{t=1}^{n} t^2 = \sum_{i=1}^{n} y_i t \end{cases}. \tag{17.4}$$

При застосуванні МНК для оцінки параметрів експонентної або логістичної функцій виникають складності з розв'язанням системи нормальних рівнянь, тому попередньо, до одержання відповідної системи, звертаються до перетворення цих функцій. Для нелінійних трендів попередньо проводять стандартну процедуру їх лінеаризації.

Іншим методом вирівнювання (згладжування) часового ряду, тобто виділення невипадкової складової, є метод **ковзних середніх**. Він ґрунтується на переході від початкових значень членів ряду до їхніх середніх значень на інтервалі часу, довжина якого визначена заздалегідь. При цьому сам обраний інтервал часу «ковзає» уздовж ряду.

Одержаний у такий спосіб ряд ковзних середніх поводиться більш гладко, ніж вихідний ряд, через усереднення відхилень ряду. Дійсно, якщо індивідуальний розкид значень члена часового ряду y_i, біля свого середнього (згладженого) значення характеризується дисперсією σ^2, то розкид середньої з m членів часового ряду $(y_1+y_2+...+y_m)/m$ біля того самого значення характеризуватиметься істотно меншою величиною дисперсії, що дорівнює $\frac{\sigma^2}{m}$. Для усереднення можна скористатися середньою арифметичною (простою та з певними вагами), медіаною та ін.

Існує кілька способів визначення типу тенденції. До числа найпоширеніших способів відносяться якісний аналіз досліджуваного процесу, побудова і візуальний аналіз графіка залежності рівнів ряду від часу. З цією самою метою можна використати і коефіцієнти автокореляції рівнів ряду. Тип тенденції можна визначити шляхом порівняння коефіцієнтів автокореляції першого порядку, розрахованих за вихідними і перетвореними рівнями ряду. Якщо часовий ряд має лінійну тенденцію, то його сусідні рівні $\hat{y}_t$ і $\hat{y}_{t-1}$ тісно корелюють. В цьому випадку коефіцієнт автокореляції першого порядку рівнів вихідного ряду повинен бути високим. Якщо часовий ряд містить нелінійну тенденцію, наприклад, у формі експоненти, то коефіцієнт автокореляції першого порядку за логарифмами рівнів вихідного ряду буде вищим за відповідний коефіцієнт, розрахований за рівнями ряду. Чим чіткіше виражена нелінійна тенденція в досліджуваному часовому ряді, тим більшою мірою будуть розрізнятися значення зазначених коефіцієнтів.

Вибір найкращого рівняння у випадку, коли ряд містить нелінійну тенденцію, можна здійснити шляхом перебору основних форм тренду, розрахунку за кожним рівнянням скорегованого коефіцієнта детермінації і середньої помилки апроксимації. Цей метод легко реалізують при комп'ютерній обробці даних.

17.3 Аналіз адитивної і мультиплікативної моделей часового ряду

Аналіз часового ряду полягає у виділенні окремих компонент. Методика аналізу залежить від того, який зв'язок між цими компонентами: адитивний або мультиплікативний.

Адитивною моделлю часового ряду називають таку модель, у якій зміна значень змінної в часі описується через додавання окремих компонентів:

$$Y = T + S + E. \tag{17.5}$$

Процедура аналізу адитивного часового ряду включає:

- розрахунок значень сезонної компоненти S;

- вирахування сезонної, компоненти з фактичних значень (десезоналізація даних), тобто Y-S;

- розрахунок тренду T на основі десезоналізованих даних;

- розрахунок помилок як різниць між фактичними й трендовими значеннями $E = Y - T$;

- розрахунок помилки апроксимації - середнього відхилення або MAD або средньоквадратичної помилки MSE.

Для виділення сезонної компоненти провадиться усунення сезонних коливань за методом ковзної середньої. На основі рівнів часового ряду розраховуються ковзні середні, які звільнені від сезонних коливань, але включають випадкову компоненту.

Наприклад, для виключення сезонних коливань з поквартальних даних знаходять усереднені «ковзні» рівні:

$$\overline{y}_3 = \frac{1/2y_1 + y_2 + y_3 + y_4 + 1/2y_5}{4}, \tag{17.6}$$

$$\overline{y}_4 = \frac{1/2y_2 + y_3 + y_4 + y_5 + 1/2y_6}{4}, \tag{17.7}$$

$$\overline{y}_5 = \frac{1/2y_3 + y_4 + y_5 + y_6 + 1/2y_7}{4} \text{ і т. д.,} \tag{17.8}$$

де $\overline{y}_i = T_i + E_i$ - елемент часового ряду, що містить тренд і випадкову компоненту.

Тоді виділення сезонної компоненти провадиться на основі рівності

$$y - \overline{y} = S + E. \tag{17.9}$$

Сезонна компонента знаходиться як середнє значення сезонних оцінок для кожного сезону незалежно від особливостей року. Загальна сума сезонних оцінок повинна дорівнювати нулю: $\Sigma S_j=0$ (тут j - номер сезону). Це необхідно, щоб усереднити значення сезонної компоненти в цілому за рік. Тому отримані сезонні оцінки доводиться коректувати, щоб виконати цю умову. Після цього усувають сезонну компоненту з фактичних даних, тобто знаходять $Y- S= T + E$. Дані такого роду називаються десезоналізованими. Їх використовують для побудови рівняння тренду.

Рівняння лінійного тренду має вигляд:

$$T = a + bt,$$

де t - номер кварталу (або місяця);

a - відрізок, що відсікається лінією тренду при перетинанні з віссю ординат;

b - характеристика нахилу лінії тренду до осі абсцис.

Параметри a і b визначають за методом найменших квадратів. Рівняння для розрахунку параметрів a і b мають вигляд:

$$b = \frac{n\sum ty - \sum t \sum y}{n\sum t^2 - \left(\sum t\right)^2}; \quad a = \frac{\sum y}{n} - \frac{b\sum t}{n},$$

де t - порядковий номер кварталу;

$y= T + E$ - десезоналізовані рівні часового ряду.

Знайшовши сезонну компоненту і тренд, виділяють випадкову компоненту:

$$Y\text{-}S\text{-}T=E. \tag{17.10}$$

Значення випадкової компоненти (помилки) використовують для розрахунку середнього абсолютного відхилення:

$$MAD = \frac{\sum_{i=1}^{n} |E_i|}{n}, \tag{17.11}$$

або средньоквадратичної помилки

$$MSE = \frac{\sum_{i=1}^{n} E_i^2}{n}. \qquad (17.12)$$

Якщо помилки малі, то роблять висновок про те, що тенденція стала і дозволяє одержати гарні короткострокові прогнози. Прогнозні значення за адитивною моделлю розраховують так

$$Y=T+S, \qquad (17.13)$$

де T - трендове значення для відповідного кварталу (місяця);

S - сезонна компонента для відповідного кварталу (місяця).

Трендове значення для прогнозного кварталу (місяця) розраховують за рівнянням тренда:

$$T = a + bt, \qquad (17.14)$$

де t - порядковий номер прогнозного кварталу (місяця).

Чим менше період попередження, тим більш обґрунтованим є прогноз.

Адитивна модель застосовується при сталості сезонної компоненти. Якщо вона зростає із зростанням тренду, то кращий результат буде отриманий на базі мультиплікативної моделі:

$$Y= T * S * E, \qquad (17.15)$$

де T - трендова компонента;

S - коефіцієнт сезонної компоненти;

E - відносний вплив випадкової компоненти.

У цьому випадку роблять вирівнювання ряду за методом ковзної середньої й знаходять коефіцієнт сезонності:

$$YY = SE. \qquad (17.16)$$

Якщо часовий ряд побудований за квартальними даними, то сума коефіцієнтів сезонності повинна дорівнювати 4, а якщо за місячними, - то $\sum_{i=1}^{n} S_i = 12$. Інакше корегують коефіцієнти сезонності, щоб виконати цю умову.

На основі десезоналізованих даних розраховують рівняння тренду - лінійне або нелінійне - і знаходять трендові значення T. Потім обчислюють помилки:

відносну $E = Y : (T*S)$,

абсолютну $E_a=Y - (T*S)$.

Близькість моделі до фактичних даних оцінюється за допомогою показників MAD і MSE.

Прогнозні значення визначаються як

$$Y = T * S, \qquad (17.17)$$

де T - прогнозне значення, знайдене з рівняння тренду.

Наприклад, $T = a + bt$, де t - порядковий номер прогнозного кварталу (місяця).

Потім розраховане значення T коректують на сезонну компоненту відповідного кварталу або місяця: $T' \bullet S$.

17.4 Спектральний аналіз часового ряду

При спектральному аналізі виходять з припущення, що часовий ряд є сумою, або спектром багатьох хвилеподібних змін, які можна описати за допомогою тригонометричних функцій. Метою спектрального аналізу є відшукання прихованих періодичностей і оцінка їх інтенсивності.

Для опису хвилеподібних коливань динамічного ряду використовують періодичну функцію Фур'є наступного вигляду:

$$\bar{y}_t = a_0 + \sum \left(a_k \cos \frac{360kt}{n} + b_k \sin \frac{360kt}{n} \right),$$

де a_0 - середній рівень ряду;

k - номер гармоніки;

i - порядковий номер часового періоду;

n - довжина ряду, тобто число рівнів у ньому.

Номер гармоніки - це те число хвиль даної довжини, які зможуть укластися в даному ряді. Наприклад, для ряду довжиною в 10 років у хвилі довжиною в 5 років номер гармоніки дорівнюватиме 2. У хвилі довжиною у два роки гармоніка дорівнюватиме 5.

Вищенаведена формула періодичної функції добре відповідає особливостям аналізу періодичних коливань в акустиці, механіці, електротехніці та в інших галузях фізики. Однак вона не підходить для економічних досліджень, де на перше місце висувається не частота коливань, а довжина хвилі l.

З огляду на те, що $l = \frac{n}{k}$, можна надати періодичній функції зручніший для застосування в економічній галузі вигляд, а саме:

$$\bar{y}_t = a_0 + \sum \left(a_l \cos \frac{360t}{l} + b_l \sin \frac{360t}{l} \right). \qquad (17.18)$$

Для знаходження параметрів (17.18) використовують формули:

$$a_0 = \frac{\sum y}{n}; \quad a_l = \frac{2}{n} \sum y \cos \frac{360t}{l}; \quad b_l = \frac{2}{n} \sum y \sin \frac{360t}{l}. \qquad (17.19)$$

При цьому дисперсію (амплітуду) хвилі l можна знайти за формулою:

$$\sigma_l^2 = \frac{a_l^2 + b_l^2}{2}, \qquad (17.20)$$

а загальну дисперсію – за формулою:

$$\sigma_{\text{çàã}}^2 = \frac{\sum (y - \bar{y})^2}{n}. \qquad (17.21)$$

Частка поясненої варіації ряду, або коефіцієнт детермінації, при цьому обчислюватиметься в такий спосіб:

$$R^2 = \frac{\sum \sigma_l^2}{\sigma_{\text{çàã}}^2}. \qquad (17.22)$$

Крім розрахунку коефіцієнтів автокореляції і складання корелограми для виявлення прихованих періодичностей використовують також періодограманаліз. Він дозволяє з певною ймовірністю визначити, скільки і якої довжини хвилі (або скільки гармонік і яких номерів) містить аналізований динамічний ряд. Виявлені хвилі можуть мати різну силу. Одні можуть бути дуже слабкими, а інші - сильними. Співвідношення сили окремих хвиль або гармонік показують наочно за допомогою графіка лінійчатого спектра Фур'є. На цьому графіку за абсцисою показують довжини хвиль, а за ординатою їх потужність. Як міру потужності цих хвиль беруть викликувані ними дисперсії, тобто

$$\sigma_l^2 = \frac{a_l^2 + b_l^2}{2}. \tag{17.23}$$

Якщо всі дисперсії практично рівні, то кажуть, що ряд характеризується «білим шумом». Наявність будь-якої періодичності в зміні рівнів ряду при цьому заперечується.

17.5 Прогнозування часового ряду

Будь-який прогноз ґрунтується на перенесенні минулих тенденцій, виявлених у ході аналізу часового ряду, на майбутнє.

Період часу, для якого робиться прогноз, називається горизонтом прогнозування. Відстань до нього від поточного періоду називається періодом попередження. Якщо період дорівнює 1-3 крокам у майбутнє, то кажуть про короткостроковий прогноз. При 4-10 кроках прогноз називають середньостроковим. При більш довгих періодах попередження говорять про довгострокові прогнози.

У ході прогнозування доводиться зустрічатися з наступним протиріччям. Довгий вихідний ряд дозволяє добре погасити всякі випадкові сплески й падіння, але створює небезпеку перенесення на майбутнє занадто старих закономірностей. Короткий ряд виключає таку небезпеку, але не захищає від впливу випадків. Прогноз не може вважатися якісним, якщо на ньому позначилися випадковості або дуже старі закономірності.

Щоб розв'язати дане протиріччя, шукають компроміс між прагненням погасити випадковості й не допустити перенесення на майбутнє занадто старих закономірностей. В умовах різких змін, характерних для нашої країни, знайти такий компроміс дуже важко: занадто короткі ряди треба використовувати, щоб не допустити присутність у прогнозі закономірностей, що втратили чинність.

Відомим виходом з такого стану може служити застосування для прогнозів так званих **адаптивних моделей**. Їх особливістю є те, що при кожному надходженні нової інформації до параметрів моделі вносять відповідні корективи. У результаті модель постійно адаптується до нових умов. При цьому систематично перевіряється близькість її розрахункових даних до фактичних рівнів ряду.

Самою складною проблемою прогнозування, що дотепер не має задовільного розв'язання, є пророкування ламання тенденції, що лежить в основі прогнозних розрахунків. Поки ламання немає, прогнозування дає мінімальні помилки. Коли воно відбувається, то самі витончені системи прогнозування дають збій і виникають великі помилки.

Аналітики біржових ситуацій на основі багаторічних спостережень за індексом Доу Джонса стверджують: імовірність ламання тенденцій виникає тоді, коли чергове падіння біржових курсів виявляється більш глибоким, ніж попереднє. Наприклад, якщо динаміка згаданого індексу за чотири періоди має вигляд: 7200, 7160, 7240 і 7180, то падіння, що відбулося наприкінці цього ряду не повинне викликати тривоги. Інша справа, якби на кінці ряду опинилася цифра менше 7160. Це було б сигналом того, що в самому найближчому майбутньому на біржі може відбутися обвал курсів акцій або цінних паперів.

Стаціонарним динамічним рядом називається такий динамічний ряд, у якого відсутня тенденція до зростання або падіння, тобто відсутній тренд.

Рівні стаціонарного динамічного ряду коливаються навколо певного середнього значення. Прогнозування зводиться до пошуку цього середнього значення.

Прогноз для періоду t+1, тобто на один крок уперед, на основі середнього рівня може мати такий вигляд;

$$П_{t+1} = (Ф_t + Ф_{t-1} + Ф_{t-2}) : 3, \tag{17.24}$$

де $Ф_t$ - фактичний рівень поточного періоду;

$Ф_{t-1}$, $Ф_{t-2}$ - рівні попередніх періодів.

Тут для розрахунку середньої взяті всього три минулих рівні. Їх може бути більше. Погашення випадковості при цьому буде сильнішим. Однак, заглиблюючись в історію, треба завжди пам'ятати про небезпеку перенесення на майбутнє занадто старих закономірностей.

Недоліком вищенаведеної середньої є те, що при її розрахунку надається однакова вага періодам, по-різному віддаленим від горизонту прогнозування. Тим часом періодам, близьким до горизонту прогнозування, треба було б давати більшу вагу. Це якоюсь мірою послабляло б вплив старих закономірностей і посилювало б значення закономірностей останніх років, близьких до горизонту прогнозування. Це особливо актуально для часу великих і різких змін в економіці, що характерно зараз для нашої країни.

Нерівні ваги можна взяти довільно. Наприклад, у такий спосіб:

Період	t	t-1	t-3
Вага	3	2	1

Тоді розрахунок середньої виглядатиме так:

$$\bar{y}_t = \frac{y_t 3 + y_{t-1} 2 + y_{t-2}}{6}.$$

Р. Браун запропонував брати ваги, що убувають за експонентою:

Період	t	t-1	t-2	t-3	…	t-n
Вага	α	$\alpha(1-\alpha)$	$\alpha(1-\alpha)^2$	$\alpha(1-\alpha)^3$	…	$\alpha(1-\alpha)^n$

Середня з такими вагами називають експонентною.

Величину α називають параметром згладжування й обчислюють за формулою, запропонованою Р. Брауном:

$$\alpha = \frac{2}{n-1}, \tag{17.25}$$

де n - число рівнів динамічного ряду, які бажано взяти до уваги при розрахунку середньої.

Якщо прогнозист визнає, що йому треба орієнтуватися тільки на чотири останніх рівні ряду, то:

$$\alpha = \frac{2}{4+1} = 0{,}4.$$

Ваги окремих періодів при цьому будуть мати такий вигляд:

Період	t	t-1	t-2	t-3	…
Вага	0,4	0,4(1-0,4)	$0{,}4(1-0{,}4)^2$	$0{,}4(1-0{,}4)^3$	…
або	0,4	0,24	0,144	0,0864	…

У зарубіжних роботах із прогнозування пропонується брати α на рівні 0,05-0,30, тобто орієнтуватися на тривалу історію: 8-40 часових періодів. Це для наших мінливих умов не підходить. Для сучасних умов України параметр α треба брати на рівні 0,7-0,9. Якщо взяти, наприклад, $\alpha = 0{,}7$, то прогноз для періоду t + 1 виглядатиме так:

$$П_{t+1} = Ф_t\, 0{,}7 + Ф_{t-1}\, 0{,}21 + Ф_{t-2}\, 0{,}063 + Ф_{t-3}\, 0{,}0189 + \ldots,$$

«хвіст» цього виразу, тобто

$$Ф_{t-1} * 0{,}21 + Ф_{t-2}\, 0{,}063 + Ф_{t-3}\, 0{,}0189 + \ldots,$$

можна представити як прогноз для періоду t. Тоді прогноз для періоду t + 1 має вигляд рекурентної формули:

$$П_{t+1} = Ф_t\, \alpha + П_t\, (1-\alpha). \qquad (17.26)$$

Розрахунки кожного чергового прогнозу є при цьому продовженням раніше зроблених розрахунків. Виключається необхідність провадити їх щораз із самого початкового періоду.

При використанні приведеної рекурентної формули для самого раннього періоду вживають «наївний прогноз» $П_1 = Ф_1$.

Можна також замість «наївного прогнозу» використати експертну оцінку. Тоді прогноз для другого періоду на базі даних першого періоду має вигляд:

$$П_2 = Ф_1\alpha + ПЕ_1(1-\alpha), \qquad (17.27)$$

де $ПЕ_1$ - прогнозна величина першого періоду, установлена експертним шляхом.

При експертній оцінці величині $ПЕ_1$ прагнуть надати таке значення, яке б у наступних розрахунках мінімізувало помилку прогнозу. Мінімізація помилки досягається не тільки перебором можливих значень згаданої величини, але також перебором різних значень параметра згладжування α. Чим менше в кінцевому результаті виходить помилка наступних прогнозів, тим краще буде адаптація моделі до реально сформованих умов.

Коли в змінах рівнів ряду можна доглянути наявність тенденції до зростання або падіння, то говорять, що ряд має тренд, тоді він є **нестаціонарним**. Наявність тренду повинна мати об'єктивне підтвердження, а не ґрунтуватися на одному тільки візуальному враженні. Для одержання згаданого підтвердження можна скористатися критерієм серій. Відповідно до цього критерію можна говорити про наявність тренду тільки тоді, коли число серій у ряді не перевищує певного критичного значення, узятого з відповідних таблиць. При цьому серією вважається послідовність елементів одного вигляду. Наприклад, послідовність елементів, менших за медіану. Або послідовність елементів, де всі вони перевищують або дорівнюють медіані.

Розглянемо приклад. Припустимо, обороти фірми за період із січня по вересень склали, млн.грн.: 10, 12, 11, 13, 13, 14, 13, 15, 13.

Чи можна вважати, що даний ряд дійсно має тренд або у ньому чисто випадкові коливання його рівнів? Для відповіді на це питання підраховуємо число серій, позначаючи рівні, менші за медіану (дорівнює 13) через А. Інші рівні - через В. Дістанемо: АААВВВВВВ. У наявності 2 серії. Критичне значення, знайдене з таблиць для 3 елементів однієї послідовності, 6 елементів другої і 5%-го рівня значущості перевірки дорівнює теж 2. Перевищення немає. Виходить, гіпотезу про наявність тренду відхилити не можна.

Зазвичай, може статися, що в даного ряду ніякої тенденції до зростання немає, а вся справа перебуває у випадковому коливанні його рівнів. Однак імовірність цього менша за 5%.

При наявності тренду прогноз здійснюють за допомогою регресійних рівнянь. Вони розглядалися нами вище як моделі динаміки. Найпростішою моделлю є рівняння прямої:

$$\overline{y}_t = at + b,$$

де t - фактор часу, тобто порядковий номер рівня ряду.

Підставивши в дане рівняння в якості t порядкові номери майбутніх періодів часу, одержимо точковий прогноз для цих періодів.

Порівняно із точковим значно більшу практичну цінність має інтервальний прогноз, що має задану імовірність. Імовірність здійснення точкового прогнозу згідно положень теорії імовірностей дорівнює нулю.

Для одержання інтервального прогнозу попередньо розраховують граничну помилку рівняння. Для її знаходження використовують формулу:

$$\Delta = t\mu \qquad (17.28)$$

де Δ - стандартна помилка (std. error of estimate);

t - квантіль розподілу Стьюдента для відповідного числа ступенів свободи, що дорівнює n-m, і заданої імовірності прогнозу (Q), що зазвичай береться на рівні 5%.

Елементарна логіка говорить про те, що помилка прогнозу не може залишатися постійною, незважаючи на збільшення періоду попередження. Вона повинна, зростати. Із збільшенням періоду попередження повинні також розширюватися межі довірчого інтервалу. Інакше кажучи, потрібне виправлення на зміну періоду попередження. Це можна здійснити за допомогою наступної формули:

$$K = \sqrt{\left(\frac{n+1}{n} + \frac{3(n+2L-1)^2}{n^3 - n}\right)},$$

де L - період попередження;

n - довжина ряду.

17.6 Зв'язний аналіз часових рядів

Вивчення зв'язку між двома рядами називають **зв'язним аналізом**. До нього прибігають тоді, коли хочуть оцінити ефективність витрат на ті або інші заходи. Наприклад, порівнюють зростання витрат на рекламу з зростанням товарообігу торгового підприємства або динаміку витрат на добрива з динамікою врожайності.

При здійсненні зв'язного аналізу треба завжди пам'ятати про можливість викривлення його результатів за рахунок впливу так званої помилкової кореляції, що може виникнути через просте супуття в часі розвитку двох явищ.

Яскравий приклад помилкової кореляції навів в одній зі своїх робіт англійський статистик Д. Фінні. Він зрівняв динаміку зареєстрованих радіоприймачів і число душевнохворих у післявоєнній Англії і одержав високий коефіцієнт кореляції, хоча прямий зв'язок тут, зазвичай, відсутній.

Супуття в часі може виникнути через наявність в часових рядів автокореляції. Під нею розуміють залежність наступних рівнів ряду від попередніх. Для характеристики автокореляції існує тест Дарбіна-Уотсона, що обчислюють за формулою:

$$DW = \frac{\sum_{i=1}^{n-1}(e_{i+1} - e_i)^2}{\sum_{i=1}^{n} e_i^2}, \tag{17.29}$$

де e_i - різниця між фактичним і вирівняним рівнем для і-го періоду.

Автокореляція відсутня і можна без небезпеки викривлень вивчати зв'язок між двома рядами, коли DW близький до 2. Автокореляція є й можливі великі викривлення рівня зв'язку між рядами за рахунок існування між ними помилкової кореляції, коли DW близький до нуля або 4.

Існують таблиці значень цього тесту. Входами в них є задана імовірність перевірки і число спостережень n. Наприклад, для $n = 15$ і 95%-ї імовірності висновків приводяться такі межі цього тесту:

- при DW < 1,08 (або DW > 2,92) існує позитивна (або негативна) автокореляція;

- при 1,08 < DW < 1,36 і при 2,64 < DW < 2,92 існує невизначеність, коли не можна з достатньою впевненістю ні відхилити, ні прийняти гіпотезу про наявність автокореляції;

- при 1,36 > DW < 2,64 автокореляція відсутня.

Наявність автокореляції не тільки утруднює вивчення зв'язку між часовими рядами, але й робить зовсім неможливим здійснення прогнозу значень рівня одного ряду за передбачуваним значенням другого ряду.

Іноді відразу видно, що ряди ніяк не можуть бути зв'язаними, тому що матеріальна природа відбиваних ними явищ не допускає цього. В інших випадках вирішити питання про наявність або відсутність зв'язку між рядами досить важко. У таких випадках потрібна перевірка рядів на наявність автокореляції і необхідне використання методів, що послаблюють або виключають вплив помилкової кореляції.

Контрольні запитання

1. Наведіть визначення часового ряду.
2. Чим відрізняються часові ряди від звичайних просторових вибірок?
3. Які фактори впливають на рівні часового ряду?
4. Які компоненти містить реальний часовий ряд?
5. Поясніть адитивну і мультиплікативну моделі часового ряду.
6. Поясніть, що таке тренд часового ряду і які види тренду зустрічаються?
7. Як визначають значущість тренду?
8. Перелічіть основні етапи аналізу часових рядів.
9. Перелічіть найпоширеніші методи аналізу часових рядів.
10. Поясніть, що таке автокореляція рівнів часового ряду. Як її можна виміряти?
11. Перелічіть властивості коефіцієнта автокореляції.
12. Що таке автокореляційна функція часового ряду і корелограма?
13. З якою метою виконують аналітичне вирівнювання часового ряду?
14. Які функції найчастіше застосовують для побудови трендів?
15. Який метод дозволяє визначити параметри тренду?
16. Які методи використовують для згладжування часового ряду?
17. В чому полягає метод ковзних середніх?
18. Поясніть, як провадиться усунення сезонних коливань за методом ковзної середньої?
19. З якою метою провадиться спектральний аналіз часового ряду?
20. Поясніть, у чому полягає зв'язний аналіз часових рядів?

СПИСОК ДЖЕРЕЛ

1. Замков О. О. Математические методы в экономике / О. О. Замков. - М.: Финансы и статистика, 2001.
2. Экономико-математические методы и модели : Учебн. пособие / Н. И. Хлод, А. В. Кузнецов, Я. Н. Жихар и др.; Под общ. ред. А. В. Кузнецова. 2-е изд. - Мн.: БГЭУ, 2000. - 412 с.
3. Экономико-математические методы и прикладные модели : Учебн. пособие для вузов / В. В. Федосеев, А. Н. Гармаш, Д. М. Дайитбегов и др.; Под ред. В. В.Федосеева. - М.: ЮНИТИ, 2001. - 391 с.
4. Вітлінський В. В. Математичне програмування: [навч. посіб для студ. вищих навч. закл. економ. спец.] / В. В. Вітлінський, С. І. Наконечний, Т. О.Терещенко. - К.: КНЕУ, 2001. – 380 с.
5. Кузнецов Ю. Н. Математическое программирование: учебник / Ю. Н. Кузнецов, В. А. Кузубов, А. В. Волощенко. - М.: Высш.школа, 1980. – 240 с.
6. Акулич И. Л. Математическое програмирование в примерах и задачах : учеб. пособ. / И. Л. Акулич .- М.: Высш. школа, 1986. – 244 с.
7. Таха Х. А. Введение в исследование операций / Х. А. Таха - М.: Изд.дом «Вильямс», 2005. – 912 с.
8. Исследование операций в экономике: Уч. пособие для вузов / Н. Ш. Кремер, Б. А. Путко, И. М. Тришин, М. Н. Фридман./ Под ред. проф. Н. Ш. Кремера. - М.: ЮНИТИ, 2003. - 407 с.
9. Акоф Р. Основы исследования операций / Р. Акоф, М. Сасиени. - М.: Мир, 1971. – 320 с.
10. Наконечный С. И. Эконометрия: учебник / С. И. Наконечный, Т. П. Терещенко. - К.: КНЕУ, 2001.
11. Кремер Н. Ш. Эконометрика : Учебник для вузов / [Н. Ш. Кремер, Б. А. Путко; под ред. проф. Н. Ш. Кремера]. - М.: ЮНИТИ-ДАНА, 2002. - 311 с.
12. Практикум по эконометрике : Учебн. пособие / [И. И. Елисеева, С. В. Курышева, Н. М. Гордиенко и др.; Под ред. И. И. Елисеевой]. - М.: Финансы и статистика, 2002. - 192 с.
13. Магнус Я. Р. Эконометрика. Начальный курс / Я. Р. Магнус, П. К. Катышев, А. А. Пересецкий. - М.: Дело, 2001. - 400 с.
14. Кулинич Е. И. Эконометрия : учеб. пособ. / Е. И. Кулинич - М.: Финансы и статистика, 2000. - 304с.
15. Афанасьев В. Н. Анализ временных рядов и прогнозирование : Учебник / В. Н. Афанасьев, М. М. Юзбашев. - М.: Финансы и статистика, 2001. – 228 с.
16. Лещинский О. Л. Эконометрия : учеб. пособ. / О. Л. Лещинский, Рязанцева В. В., О. О. Юнькова. - К.: МАУП, 2003.
17. Шалабанов А. К. Эконометрика : Учебно-методическое пособие / А. К. Шалабанов, Д. А. Роганов – Казань: ТИСБИ, 2002. – 56 с.
18. Доугерти К. Введение в эконометрику : Пер. с англ. / К. Доугерти. - М.: Финансы и статистика, 1999. - 402 с.

19. Вітлинський В. В. Аналіз, оцінка і моделювання економічного ризику. - К.: «Деміур», 1996. – 212 с.

20. Вітлинсъкий В. В. Ризик у менеджменті / В. В. Вітлинсъкий, С. І. Наконечний. - К.: ТОВ «Борисфен-М», 1996. - 326 с.

21. Вітлінський В. В. Аналіз, моделювання та управління економічним ризиком : Навч.-метод. посібник / В. В. Вітлінський, П. І. Верченко – К.: КНЕУ, 2000.

22. Гранатуров В. М. Экономический риск: сущность, методы измерения, пути снижения : Уч. пособие / В. М. Гранатуров. – М.: Дело и Сервис, 1999.

23. Світлична Т. І. Теоретичні основи курсу «Економічний ризик та методи його вимірювання» : Навч. посіб. / Т. І. Світлична. – Харків: ХДАМГ, 2000.

24. http://www.mathnet.ru.

25. http://www.aup.ru/articles/investment/1.htm.

Навчальне видання

ВОРОНКОВ Олексій Олександрович
МЕДВЕДЄВ Ігор Анатолійович

КОНСПЕКТ ЛЕКЦІЙ
з курсу

ЕКОНОМІКО-МАТЕМАТИЧНІ МЕТОДИ І МОДЕЛІ

(для слухачів другої вищої освіти за спеціальностями 7.03050401 - Економіка підприємства, 7.03050901 - Облік і аудит)

Відповідальний за випуск *Т. А. Пушкар*

За авторською редакцією

Комп’ютерне верстання *І. В. Волосожарова*

План 2015, поз. 184 Л

Підп. до друку 03.11.2015 Формат 60x84/16
Друк на різографі Ум. друк. арк. 11,7
Тираж 50 пр. Зам. №

Видавець і виготовлювач:
Харківський національний університет
міського господарства імені О. М. Бекетова,
вул. Революції, 12, Харків, 61002
Електронна адреса: rectorat@kname.edu.ua
Свідоцтво суб’єкта видавничої справи:
ДК № 4705 від 28.03.2014 р.